

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

SUMMARIES

OM nucleic - nucleic search, using sw model

Run on: April 4, 2003, 19:23:52 ; Search time 7537 Seconds
(without alignments)
9657.174 Million cell updates/sec

Title: US-09-954-950-1

Perfect score: 2501

Sequence: 1 cggcagcagatttcagtc.....aaaaaaaaaaaaaaaaaaaaa 2501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_da:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_da:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htg_mus:*

34: em_htg_pln:*

35: em_htg_rtd:*

36: em_htg_mam:*

37: em_htg_vrt:*

38: em_sy:*

39: em_htgo_hum:*

40: em_htgo_mus:*

41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	2501	100.0	2501	6	AX461660 Sequence
2	926.2	37.0	2316	8	AF360278 Arabidops
3	926.2	37.0	2381	6	AX461662 Sequence
4	926.2	37.0	2381	8	AX461662 Sequence
5	540.8	21.6	160704	8	AP003238 Oryza sat
6	321.4	12.9	2503	6	HSMLHI
7	319.8	12.8	2484	6	AR122159 Sequence
8	319.8	12.8	2484	6	AR131133 Sequence
9	319.8	12.8	2484	6	AX214172 Sequence
10	319.8	12.8	2484	6	AX234588 Sequence
11	319.8	12.8	2484	6	HS007343 Human DNA m
12	319.8	12.8	2530	9	BC006850 Homo sapi
13	315.2	12.6	2068	3	AY069160 Drosophill
14	300.6	12.0	2548	10	BC021815 Mus muscu
15	299	12.0	2283	10	AR250844 Mus muscu
16	291	11.6	2274	10	RN080054 Rattus norv
17	261.4	10.5	2385	6	AR131262 Sequence
18	249.8	10.0	132171	3	AC008370 Drosophill
19	249.8	10.0	157559	2	AC020344 Drosophill
20	249.8	10.0	226833	3	AE003838 Drosophill
21	247.2	9.9	2233	3	AF068257 Drosophill
22	212.8	8.5	3218	6	AX214167 Sequence
23	212.8	8.5	3218	6	AX234583 Sequence
24	212.8	8.5	3218	8	SCMLHI
25	208	8.3	41200	8	SC8520X
26	185.6	7.4	891	11	CMS06KLN
27	141.6	5.7	727	6	A86098 Sequence 75
28	141.6	5.7	727	6	AR155591 Sequence
29	141.6	5.7	727	6	AR155591 Sequence
30	135.4	5.4	2664	3	AF346620 Trypanoso
31	133	5.3	11613	1	AE005650 Escherich
32	133	5.3	243184	1	AP002568 Escherich
33	128.6	5.1	2446	1	STYMUTL
34	128.6	5.1	20418	1	AE008904 Salmonell
35	128	5.1	10151	1	AE013665 Yersinia
36	128	5.1	214050	1	AJ414142 Yersinia
37	127	5.0	249050	1	AL627283 Salmonell
38	125	5.0	1848	1	AF440199 Escherich
39	125	5.0	2300	1	ECMUTIG
40	125	5.0	12088	1	AE000489 Escherich
41	125	5.0	338534	1	ECODW93 U14003 Escherichia
42	122.6	4.9	13668	1	AE004907 Pseudomon
43	117.4	4.7	2308	9	BC005833 Homo sapi
44	117.4	4.7	2308	9	BC005866 Homo sapi
45	116.6	4.7	13571	1	AE004982 Halodacte

ALIGNMENTS

RESULT 1

AX461660

LOCUS AX461660 2501 bp DNA linear PAT 08-JUL-2002

DEFINITION Sequence 1 from Patent WO0224890.

ACCESSION AX461660

VERSION AX461660.1 GI:21726826

KEYWORDS

SOURCE

ORGANISM

Oryza sativa.

Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE

1 Mahajan, P. B.
Rice mhl ortholog and uses thereof

Pred. No. is the number of results predicted by chance to have a

Query Match	Similarity	65.4%	Score	926.2	DB	8	Length	2316
Best Local	Conservative	0	Mismatches	733	Indels	12	Gaps	3
158	GGGAGCCGCCCGCCGATCCGAGGTTGAGGAGTCCGGTGGTGAACCGCATCCGCGCGGGG	217						
159	GGAGGCCGCCAGAGATTCACGCTTAGAAGAAATCAGTACAGCCGTAATCCGACGCTGGTG	116						
218	AGGTATTCACAGCGCGCTGCTGCGGGGGAAGGAGCTCATGAGAAACAGCCCTGACGCGT	277						
117	AAGTATTCACGCGCTGCTGCTGCGGGGGAAGGAGCTCATGAGAAACAGCCCTGACGCGT	176						
278	GGCGCTCCAGGCGTCCGCTTGGCGGGAAGGAGCGGTGGCCTCAAGCTCATCCAGTCTCG	337						
177	ATTCAGTTCCATTAAGCGTCCGTTGTCATTAAGAGAGCGGTGTTGAACCTCATTTCAAGTCTCG	236						
338	ATGACGCCCATGCGATCATGAGTTTAGAGATTTGGCAATATTTGGCCGAAGGCTATCTACT	397						
237	ACGACGGTCACGCTATTTAGACGCTGAAGCTTCCCATACTATGACGAGACATACAAAT	296						
398	CAAGTTATTCGATACGAGGATTCGACAGACCATTAATTAATCGATGGGGTTCAGAGGGAGG	457						
297	CGAAGCTACCTAGTTTGGAGATTTGTTCTCTCTGAGTTCAATGGGATTTAGAGAGAG	356						
458	CTTGGCTAGTATGACTATATTTGGCCATATTTACCGTGCAACGATTAACAGAGGCCAAT	517						
357	CATTAGTAGTATGACCTATGTTGCTCATGTTACAGTGAAGTACTACTATTAATAAGGCCA	416						
518	TGCAAGGCTACAGGCTTCTTACAGAGATGGGATATGAGATGGAGTACGCTTAAGCCCTTGG	577						
417	TTCAAGGTTATAGAGTCTTATAGAGATGGTGTCTAGAGAGCATGACCAACAAAGGCTGTG	476						
578	CTGCGGTGAAGGAACTCAAGTATGTTGGAATAATCTATTTACAACTATGTAAGCCCGA	637						
477	CTGCTGTCAAGGAAACAGATTAATGGTGAAGATTTGTTACAAATATGATTTGTAGAA	536						
638	AGAAACATTCAGAACTCCATATGATGATTAACCCAAAGATCGTAGACTCATACAGTCGGT	697						
537	GGAAACACTTCAAAATTTGCTGATGATTAAGGAAATCGGGAATTTGCTTAAGCCGA	596						
698	TTGCAAGTCCATCAATCAAGTATGTTTACCTTCTCTGCAAGAAAGCATGAGCCATAGAGCAG	757						
597	TGGCATTCATTAACATTAATGTCAGCTTTCTGTCGAAAGCATGAGCGTTAAAGCGT	656						
758	ATGTTCAATAGTGCAGTACATCTCTCAAGTTAGATCTATACAGAGTGTATGGGCTT	817						
657	ATGTTCAATAGTGCAGTACATCTCTCAAGTTAGATCTATACAGAGTGTATGGGCTT	716						
818	CTGTGCTTGTATCTCATAGAAATTAAGGTTTCATATGAGATGGTGTGAGATTCATCT	877						
717	CAGTTGCAAGAACTGATGAAAGTGAAGTTCCTCTGCTGACTCTCTGCTGCTGCTGCT	776						
878	TCAGATGATGTTATCATCTCAATGCAATTAATTTGGCAAGAAAGATTAACATATTC	937						
777	TGATATGAGAGGTTTCATATCCAAATTCATCACTATGTTGTAAGAAGACATATTTGGTGC	836						
938	TTTTCAATTAATATAGGCTTGTAGATGATGATCTTGAAGAAAGCATATGAAATTTGTG	997						
837	TTTTCAATTAATATAGGCTTGTAGATGATGATCTTGAAGAAAGCATATGAAATTTGTG	896						
998	ACTGTGCAACATTCCTCAAGCATGCAACCATTTTCATATACATGTCATACATCTTCAT	1057						
897	ATGCTCAACATTCCTCAAGCATGCAACCATTTTCATATACATGTCATACATCTTCAT	956						
1058	CAGAACAGTGTATTAATATACACCCAAAGGAAAGGTTAGGCTTTTGAATCAAG	1117						
957	GGGAACATGTTGATATCAATATTCACCCAAAGGAAAGGTTAGGCTTTTGAATCAAG	1016						
1118	AGGTTATTTATGAACATTAAGAAATGCTTATGAGAAAGGATGATGATTTATACAA	1177						
1017	AAATCATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1076						
1178	CAAGATATTCCAACTCAGGCAATTAACCTTATACAGGATTTGCTCAAGCTTAACCCAA	1237						

Db	1077	CTAGACAGCTTTTCAAGACGAGAAAGTGGAAATACATTCATCTACGTTTAAACATCTCGAAAA	1136
Qy	1238	AGGATTAAGTTTCTGAGGGCAATGATGGGTTCTGGAAACAATAATCTCAAAAATTCCTGTGA	1297
Db	1337	GTGAT---TCTCAGTTTTCTCAGAAAGCCTTTCTGGACAAAAGACACAGAAAGTTCTCTGTGA	1193
Qy	1298	GCCAAATGCTCAGAACAGATCCACGGCAATCCATCTGGAAAGATTGCACACTACTGGCAGC	1357
Db	1194	ACAAATATGTGAGAACAGATTCATCTAGATCCAGCTGGAAAGGTTACATCCCTTTTGGAC	1253
Qy	1358	GGCAATCTTAAATCT---TGAATAACAATTTGATCTGTATCTGTAGAAATATGTGTAA	1414
Db	1254	CCAAAGCCAAAGAGCTCTCCCGAAGAGGTTTCTAGTTTGAGTGTAGAGGCTTCTGTAA	1313
Qy	1415	GATCAAGAGAAACCAAAAAGATGCTGTGATTTGTCAAGCCGTCATGAGCTCTGTG	1474
Db	1314	GGCAAAAGAAAGAACCCAAAGAAACGTGATCTTTATGATGTCCAGAACTTATGTCTG	1373
Qy	1475	AAATAGATTCTAGCTTCATCTGTGGCCTTTTGGACATTGTCAAGACCTGCACATATGTTG	1534
Db	1374	GAGTTGACAGCTGCTGCATCCAGATATCTGTGAGACGTGAAGAAATTCACAAATGTG	1433
Qy	1535	GACTTGGCGAAGACCTTTGGCTTTGATTAACAACAATATCCCGTTATACCTGTAATG	1594
Db	1434	GAATGGCAGATGATGTTTGTGCTTATGTTCAAGTATACACCACTATATATACCAATG	1493
Qy	1595	TGCTAAATATTAGTAAAGAACTTATGTACACAGACACTTTGTGCGTTTGGAGACTTCA	1654
Db	1494	TGTGAATCTCAGCAAGACCTAATGATACAGAAACTCTGTGTGTTTCTCATTTTA	1553
Qy	1655	ATGCTATTCAGCTCAGTGAACCAAGCTCCACTTCAGAGTTGGTGGTAAAGGACAGTGAAG	1714
Db	1554	ACGCAATACACTTTAGGATGCAGCCCTTTGTCAAGATGTGAATGTTGGCTCTGAAG	1613
Qy	1715	ACGATGAAT-----GATAGATGAAAGAAAGGATATAGAAGACTGAGATTCGAGAAG	1768
Db	1614	AGGAGATCTGATTCAGGAAATGATACAAAAGATGATCTGAAACAAAGAAATTCGTGAAA	1673
Qy	1768	TAAACCTGAGATCTTAAAGAAAATGCTGAGATGATTAATAGATCTTTCTATTCACA	1828
Db	1674	TGAATACAAAGACTCTCAGAGAAACCAAAAATGTTAAGAGATTAATTCACAGCTGCACA	1733
Qy	1829	TTGATCAAGATGGGCAAAATGTGACAGACTTCCTGTTACTGGACAGTACACCCCTGATA	1888
Db	1734	TTGACTCCAGTGCAAAATTTGTCAAGGCTTCTGTGATCTCGACAGATACACTGTACA	1793
Qy	1889	TGGACCGCTTCCAGAAATTTGTGTGGCTTTAGGAATATGTTACTGTTGGATGTGACGGA	1948
Db	1794	TGGATCCGTTCTCGAATTTTTTACTATGCTTGGAAATATGTTGATGGGAATGAGA	1853
Qy	1949	AAGATGTGCTCAGAACAGTACTCTCTGTAGGAACCTTATGCACTTCAATCCCCAA	2008
Db	1854	AAGATGTGCTTTCAGAGATTTCTCAGACTATGCGGAACCTTTTACGCAATGATCCTCTC	1913
Qy	2009	TCCTTCCAATTCATCTGGGAATGGATTCATTTATPACAGAAAAATAGAAATTAATGG	2068
Db	1914	TTTTCCCAAAACCAATCGGGTACGATTTTCAAGTTCTATAGTAAGAGGTGAGACTCTC	1973
Qy	2069	CTGATGAACATGCTGAGAAATGATGTATATTCAGATGAAATATGCTTATTCAGAAACTTC	2128
Db	1974	AGGAAAGTCAGATTTAGAGGGTAACTCGATATGAGAGCAATCTTTCACCAAGATCTTC	2033
Qy	2129	TTGGGGAAGCAGAAACAGACATGGGCCCAACGTAGATGGACATTCAGATGCTTGTTC	2188
Db	2034	TGTGAGATGCTGAAACGCAATGGGCGACAAAGTGAATGCTCAATCCAAACAGTGTGTTTC	2093
Qy	2189	CATCATGAGATTTTCTCTCAAGGCCCGCGGAAGCAATGCGAACAAGATGAAAGTTTGGC	2248
Db	2094	CGTCATATGATGTTCTTTGAAGCCACCAAGCTTCACATGAGCTTTAAATGGGACTTTGTGA	2153
Qy	2249	AGGTTGCTTCCCTGGAGAACTCTACAAATTTTTGAAGAGTTTATGCTCATATA 2301	

Db 2154 AGGTACGATCCCTTGAAAGCTGTACAGATATTCCAGACGATGCTAAGTGA 2206

RESULT 3
AX461662 2381 bp DNA linear PAT 08-JUL-2002
LOCUS Sequence 3 from Patent WO0224890.
DEFINITION AX461662
ACCESSION AX461662
VERSION AX461662.1 GI:21726828
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE
1 Mahajan, P. B.
Rice mhl ortholog and uses thereof
Patent: WO 0224890-A 3 28-MAR-2002;
PIONEER HI-BRED INTERNATIONAL, INC. (US)

FEATURES
source
1. location/Qualifiers
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
3.2216
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD38486.1"
/db_xref="GI:21726829"
/translation="MIDSSLAEMEEREPATTIVPREPKIORLESVVRNIAAG
VTORVSAKELVENSIDSSISVYKDGKLIQVDDHGIRREDLPILCEHRT
TSKLRFEDELSSWGRGEBALSHVYAHVTVTITQGLHGKRVSRDMEHEP
RACAAVKGQIWMENLFYMIARRTIOLASADYKGLVDLSRMALHYNVEFSKXK
GAVKADVHSVGSRLDSIRSYGVSAKMLKVEVSCDSGCEDEMGFTSNSKY
AKRTILVEINRLEVCALKRAIEIVYATLPKSPVYSINLPREHVDINIHPT
KEVSLINELIEMIOSVEYKLRNANDTFROKXVYIOTSLTSOKSDSPVYQKP
SGQOKVPMKAVRTDSDPAGRLHAFIQRPQSPDKVSSLSVRSVRSRPNKE
TADLSVDELIVGDSCHPGMLFETVRYGMADVVALVQYNTHLILANVNSK
ELMYOOLRFHFAPNALISDPAISELILALKEEDLPDNDTKDILERLANVNE
LKERAEMLVEEYFVHIDSANILSRPLVLDQYTPMDVPEFLICIGDVEDEKRS
CEGVSAAGNFYAMHPPLPNPSSGDIQFYKSGSSQSEKDLBENVMEINLDDLI
LSDANAMQRMSTIOHVLFPSPMLFIRKPMASMSNGTVKVASLEKLIKIRERC"

BASE COUNT 721 a 450 c 548 g 662 t
ORIGIN

Query Match 37.0%; Score 926.2; DB 6; Length 2381;
Best Local Similarity 65.4%; Pred. No. 2.2e-178;
Matches 1408; Conservative 0; Mismatches 733; Indels 12; Gaps 3;

QY 158 GGGAGCCGCCCGCATCCGAGGTGGAGAGTGGGTGTGAACCCGATCGCGGGGG 217
Db 73 GAGAGCCACCGAAGATTCACGCTTAGAAGAAATCAGTAGCAACCGATCGAGCTGTG 132
QY 218 AGGTATTCACGCGCGGTGGGTGGAGAGTGCATCAAGCAACAGCGCTGAGCGTGT 277
Db 133 AAGTAATTCACGCGCTTCAAGTTTCAGCTGTGAAGAGCTCGTTGAGAACAGCGCTGAGCGCG 192
QY 278 GCGCCTCCACAGCTCTCGCTGTGGGTGAAGACGCGCTCAAGCTCATCCAGGTCTCCG 337
Db 193 ATTCAAGTTTCATTAAGCGTGTGTCAAGAGCGGTGTGAAGATCATTCAGATCTCCG 252
QY 338 ATACAGCGCATTCAGATGTTGAGAGTTCGCAATATTGTGCGAAGAGCATTAAGT 397
Db 253 ACAGCGCTCAGGTATTAGACGTGAAGACTTGCAGTACTATGCGAGAGCATACAT 312
QY 398 CAAGATTCATGATACGAGATCTGCAGACATTAAGATGAGGGGTGAGAGGGAGG 457
Db 313 CGAAGCTGAAGTTCAGAGATTTGTTCTCTGAGTTCAATGAGGATTAAGAGGAGG 372
QY 458 CTTCGCTAGTATGACTTATGTTGCGCATGTACCGTGACAGACATTAAGAGAGCCAT 517
Db 373 CATTAAGTATGATGACTATGTTGCTCATCTTTACAGTACATTAAGAGCCAGA 432

QY 518 TGCAGCGCTACAGCGTTTCTTACAGAGATGCTGTATGAGAAATGACCTTAAGCTTCG 577
Db 433 TTCTAGTTTATGAGTGTCTTATAGAGATGTCATAGAGATGAGCAACCAAGCGCTGTG 492
QY 578 CTGCGGGAAGAGCAACGATGATGTTGAAAATCATTTTCAACATAGTGAGCCCGCA 637
Db 493 CTGCTGCAAGAGAACACATATATGTTGGAATTTGTTCTCAAAATATGATGCTAGAA 552
QY 638 AGAAACATTCAGAACTCCAAATGATGATACCCAGATGCTAGACTTATGATGCTAG 697
Db 553 GGAAGACACTTCAAAATTCGTGATGATATTCGGAAGAAATTCGTGATTTGCTAGCCGGA 612
QY 698 TTGAGTCCATTCATACATACAGCTTACTCTTCTGAGAAAGATGAGCAATAGACAG 757
Db 613 TGCTATTCATTCATTAATTAAGTACAGCTTCTTCTGCAAGATGAGAGCTGTAAAGCTG 672
QY 758 ATGTTATAGTGAAGTACATCTCAAGCTTATAGTATGATGATGAGAGTATAGGGCTT 817
Db 673 ATGTTACATGATGCTGCTCAAGCTTCAAGCTTGTATTCATATGATGATGATGAT 732
QY 818 CTGCTGCTGATCTCATGAATAAAGTTTCATATGAGATGCTGAGATTCATCT 877
Db 733 CAGTTGCAAGAACTGATGATGAAGTATGAGTTTCTCTGACTCTCTGCTGTTGATCT 792
QY 878 TCAAGATGATGCTTACATCTCAATGCAATATATGCGCAAGAGATTAATGATATTC 937
Db 793 TTATATGAGAGGTTTATTCATTCATTAATCTAATATGTTGTAAGACATATATGTTG 852
QY 938 TTTTCATTAATGATGAGCTTGTAGACTGTACTGTTTGAAGAGATTAATGATTTGT 997
Db 853 TTTTCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 912
QY 998 ACTCTGCAATGCTCCCAAGCAATCCCAACCTTTCATATGATGATGATGATGATGAT 1057
Db 913 ATGCTGCAATGCTCCCAAGCAATCCCAACCTTTCATATGATGATGATGATGATGAT 972
QY 1058 CAGAACGCTGATGTTATATACCCCAACCAAGAGAGTTAGCTTTGATCAAG 1117
Db 973 GGGAACTGTTGATATCAAAATTCACCAACCAAGAGAGTTAGCTTTGATCAAG 1032
QY 1118 AGGTATTCAGAAATTAAGAAATTCATGAGAAATTCATGATTAATCA 1177
Db 1033 AAATCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1092
QY 1178 CCAAGATTCACAACTCAGAGCATTAATCAATGAGGATTCAGATCAACCAACA 1237
Db 1093 CTGAGAGCTTCAAGAGCAAAATGATGATCAATCAATGATCAATCAATCAATCA 1152
QY 1238 AGATTAAGTTCAGAGCATGATGATGATGATGATGATGATGATGATGATGATGAT 1297
Db 1153 GTGAT--TCTCAGATTCAGAGAGCTTCTGACAAAGACACAAAGATTTCTCTGA 1209
QY 1298 GCCAAATGTCAGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1357
Db 1210 ACAAAATGTCAGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1269
QY 1358 GGCATTCCTCAAACT--TGAAAGAAATTTGATCTGTATGTAAGAAATGTTGTA 1414
Db 1270 CCAAGCCACAAAGTCTCCAGCAAGTTCTAGTTGAGATGATGATGATGATGATGAT 1329
QY 1415 GATCAAGAGAAACCAAAAGATGCTGATGATGATGATGATGATGATGATGATGATGAT 1474
Db 1330 GGCAAAGAAACCAAAAGATGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1389
QY 1475 AAATGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1534
Db 1390 GAGTTGACAGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1449
QY 1535 GACTTCCGATGAGAGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1594
Db 1450 GAATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1509
QY 1595 TGCTAAATTAATTAAGAACTATGATGATGATGATGATGATGATGATGATGATGATGAT 1654

QY 758 ATGTCATAGTCAGATACATCTCAAGGTTAGATGCTATCAGAGATGCTCTANGGGCTT 817
 DB 673 ATGTCATAGTCAGATACATCTCAAGGTTAGATGCTATCAGAGATGCTCTANGGGCTT 732
 QY 818 CTGTCCTGCTGATCTATAGAAATTAAGGTTTATGAGAGATGCTGAGATTCATCT 877
 DB 733 CAGTTCGAAAGAACTGATTAAGATAGAGTTCCTCTGAGCTCTCTGCTGCTT 792
 QY 878 TCAAGATGATGTTATCATCTCAATATGATGCAAGAAAGATTAATCATATGATTC 937
 DB 793 TTGATATGAGAGGTTTATCATCTCAATATGATGCAAGAAAGATTAATCATATGATTC 852
 QY 938 TTTTCATTAATGATAGGCTTGTAGATCTGATCTGCTTTGAAAAGAGCTATTAATTTGT 997
 DB 853 TTTTCATTAATGATAGGCTTGTAGATCTGATCTGCTTTGAAAAGAGCTATTAATTTGT 912
 QY 998 ACCTGCAATGTCCTGCAAGATCCAAACCTTTCATATCATCTCATATCTTCTCAT 1057
 DB 913 ATGTCGCAATGTCCTGCAAGATCCAAACCTTTCATATCATCTCATATCTTCTCAT 972
 QY 1058 CAGAAACATGATGTTATTAATACACCAACCAAGAAAGATGAGCTTTTGAATCAAG 1117
 DB 973 GGGAAACATGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1032
 QY 1118 AGCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1177
 DB 1033 AAATCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1092
 QY 1178 CAGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1237
 DB 1093 CTAGGACGTTTCAAGACCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTA 1152
 QY 1238 AGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1297
 DB 1153 GTGAT---TCTCCAGTTTCTCAGAAAGCTTCTGCAACAAAGACACAGAAATTCCTGTA 1209
 QY 1298 GCCAATGTCGACAGATTCACGCAATCCATCTGGAAGATTCGACACCTTCTGCGACG 1357
 DB 1210 ACATAATGTCGACAGATTCACGCAATCCATCTGGAAGATTCGACACCTTCTGCGACG 1269
 QY 1358 GGCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1414
 DB 1270 CCAAGCCACAAAGTCTCTCCAGCAAGGTTCTTCTGATTAAGTGAAGGCTTCTGTA 1339
 QY 1415 GATCAAGGAGAAACCAAAAGATGCTGATTTGTCAAGCCGTCATGAGCTCTTGTG 1474
 DB 1330 GGCATAAGAAACCAAAAGATGCTGATTTGTCTAGTCTCAGAGAACTTATTTGCTG 1389
 QY 1475 AAATGATTTAGCTTCATCTGCTGCTTTTGGACATTCGACAGAACTGCAATATGTTG 1534
 DB 1390 GAGTTCGACGCTGCTGACATCCAGGATGCTGAGAGCTGTAAGGAATTCACATATGTTG 1449
 QY 1535 GACTTCGCGATGAAGGCTTGTGATTAACACATTAACCCGCTTAACTCTTAATG 1594
 DB 1450 GAATGCGAGATGATGTTTGTCTTACTTCAATTAACACATTAACCCGCTTAACTCTTAATG 1509
 QY 1595 TGGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1654
 DB 1510 TGGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1569
 QY 1655 ATGCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1714
 DB 1570 ACAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1629
 QY 1715 ACAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1768
 DB 1630 AGGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1689
 QY 1769 TAAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1828
 DB 1690 TGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1749

QY 1829 TTGATCAAGATGCAAAATTAAGCAAGACTTCTGTTTACTGAGACCACTACACCCCTGATA 1888
 DB 1750 TTGATCAAGATGCAAAATTAAGCAAGACTTCTGTTTACTGAGACCACTACACCCCTGATA 1809
 QY 1889 TGGACGCTTCTCCAGATTTTGTGTGCTTATGAGAAATGATTTACTTGGATATACAGA 1948
 DB 1910 TGGATGCTTCTCCAGATTTTGTGTGCTTATGAGAAATGATTTACTTGGATATACAGA 1869
 QY 1949 AAGATGCTTCTCCAGATTTTGTGTGCTTATGAGAAATGATTTACTTGGATATACAGA 2008
 DB 1870 AAGATGCTTCTCCAGATTTTGTGTGCTTATGAGAAATGATTTACTTGGATATACAGA 1929
 QY 2009 TCTTCCAAATTCATCTGCGAATGATTCATTTATACAGAAATAATGATTAATCAATG 2068
 DB 1930 TTTTCCAAATTCATCTGCGAATGATTCATTTATACAGAAATAATGATTAATCAATG 1989
 QY 2069 CTGATGAACATGCTGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2128
 DB 1990 AGGAAAGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2049
 QY 2129 TTGCGAAGCAGAAAGCAGATGAGGCCCCCAAGCTGATGAGACCATTCATGCTTCTTTC 2188
 DB 2050 TGTGAGATGCTGAAAGAGCAGATGAGGCCCCCAAGCTGATGAGACCATTCATGCTTCTTTC 2109
 QY 2189 CATCATGCGACTTCTCTCAAGCCCGGAGATGATGAGCAAGATGAGCAAGCTTGTGC 2248
 DB 2110 CGTCAATGAGATTTGTTTGAAGCCACAGCTTCCATGCTTCAATGAGCAAGCTTGTGA 2169
 QY 2249 AGTTCGCTCTTGGAGAAACTCTACAAAGATTTTGAAGGCTTACTCTATA 2301
 DB 2170 AGTTCGCTCTTGGAGAAACTCTACAAAGATTTTGAAGGCTTACTCTATA 2222

RESULT 5
 AP003238
 LOCUS 160704 bp DNA linear PLN 13-APR-2002
 DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
 PAC clone: P0401G10.
 ACCESSION AP003238
 VERSION AP003238.4
 KEYWORDS GI:20146211
 SOURCE Oryza sativa (japonica cultivar-group) (cultivar: Nipponbare) DNA,
 clone: P0401G10.
 ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1
 Sasaki, T., Matsumoto, T. and Yamamoto, K.
 Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
 PAC clone: P0401G10
 Published only in Database (2001)
 JOURNAL 2 (bases 1 to 160704)
 REFERENCE AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
 TITLE Direct Submission
 JOURNAL Submitted (19-FEB-2001) Takuji Sasaki, National Institute of
 Agrobiological Sciences, Rice Genome Research Program, Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail: tsasaki@nias.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
 Tel: 81-298-38-7441, Fax: 81-298-38-7468)

COMMENT
 On Apr 12, 2002 this sequence version replaced gi:17933034.
 Genes were predicted from the integrated results of the following:
 GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor
 (October 1998 version). The genomic sequence was searched against
 NCBI Nonredundant Protein database, nr
 (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
 RCP. Protein homologies of the coding regions were searched against
 NCBI Nonredundant Protein database with BLASTP2.0. ESTs represent
 the identified cDNA sequences using BLASTN 2.0 with the
 corresponding DBJ accession no. and RCP clone ID.

A gene with identity or significant homology to a protein is
 classified based on the protein name to indicate the homology level
 such as same name, 'putative-' and '-like protein'. A gene without

significant homology to any protein but with EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted with a gene prediction program is classified as a 'hypothetical' protein.

The orientation of the sequence is from T7 to SP6 of the PAC clone. This sequence of P0401G10 clone has an overlap with O1294.F06 clone (DBJ: AP004326) at 5' end and an overlap with P0483G10 clone (DBJ: AP003263) at the position 112,620 to 160,704 of 3' end. The sequence of this clone ends at the position 48,085 of P0483G10. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://ftp.dna.affrc.go.jp/Genomeseq.html>.

FEATURES

source
1. 160704
/organism="Oryza sativa (japonica cultivar group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="1"
/clone="P0401G10"
join(3181..4553,4587..4623)
/gene="P0401G10.1"
join(3181..4553,4587..4623)
/gene="P0401G10.1"
/note="contains ESTs"
C26818(C50145), C22453(C62707), C22454(C50145),
C96934(C62707)
/codon_start=1
/product="protein kinase-like"
/protein_id="BAB88994.1"
/db_xref="GI:20146212"
/translation="MRHALLGVFALCLLRRAOQPYDPTAKPSTTWANTDAALHH
VATYTSVARALLRLRLPARLGGSPFAFGFECNHGACAPCEFLGVAVVYCNAGSI
TAVTGTPOVVSANRABAVGCAATRELTADIDYLRSPGGVLMASAGAPGSGVMS
INDSGALVLEDSNRTWQSFDPIDTILVAGSLTQCARLTANAFDMSRRTYLAIA
ADGGLAAYDAKPPQRYVYLGSKNAGAAATNGSLAYLDRPGGQALATQVLA
GTYOYRLEHSDHLRLYEWSNGMREAGDVLRHYPGDCAPVYCGAIVGCTDMS
CPYANFRADVDFRRPGCVPTSPAPATCSRAHRLVSLRDTAFNSDITMRT
LRRVGAACACACAGCAGCAAAAFYGVGDPNGFCYLGSEVLTLMQPEVYHNSMS
HKIAGCSILYIKKESK"
complement(join(5066..5279,5401..5468))
/gene="P0401G10.2"
complement(join(5066..5279,5401..5468))
/gene="P0401G10.2"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB88995.1"
/db_xref="GI:20146213"
/translation="MNFSLHYGTYTVAKGAREELMOLFNRPSGNOESHSAGRGIS
AAEERRRRRGRGSRSGDARGTOTRRRRVPSDDMTGHRPKS"
join(6174..6277,6665..6986,7334..7735,7878..8222,
8572..8676)
/gene="P0401G10.3"
join(6174..6277,6665..6986,7334..7735,7878..8222,
8572..8676)
/gene="P0401G10.3"
/note="hypothetical protein
similar to Arabidopsis thaliana chromosome2, At2g28480"
/codon_start=1
/protein_id="BAB88996.1"
/db_xref="GI:20146214"
/translation="MATRVLSQNRKLKASLINTSORAPISPEPLRPAISFGIC
LNPPYQSPPMVAMASHSVNLVLDGPKREIEVEBSKRGITTKRLQRR
EKRRKANKNDPRIRIKGKIKOKPTEARAKTITERAKLEAMLYLAKYEIA
KAQCPGAKODLDEERFYLKYSOKSNVTPRGVFGVILNMLHMKHEITK
ICPKCPGOIOEVASEIARLSGIPINTIGNDIIVFYRGKNVOPDVMSVDTLSKK
ALESKYESLTYRRIAYSEKELIYHIALYGNOSONDYPCDDRRASLKM
EEPSQGLDSHMDNGSDITDASDEENPSYDNDDETGNIGTVLYDGGFRL
IVHAKTSVGRYSOLKLVVAGEOYHEFKLESA
join(12510..12707,13071..13212,14284..14395,16176..16212,
16496..16849)
/gene="P0401G10.4"
join(12510..12707,13071..13212,14284..14395,16176..16212,
16496..16849)
/gene="P0401G10.4"
/note="DNA mismatch repair protein"
/codon_start=1

misc_feature

gene
complement(19970..23375)
/gene="P0401G10.5"
complement(19970..23375)
/gene="P0401G10.5"
/note="probably inactive due to frame shift(s) in CDS
pseudogene, similar to GDSL-motif lipase/acylhydrolase"
/pseudo
join(25336..25599,25803..26041,26224..26428,26676..26744,
26826..27012,27717..27883,28412..29002)
/gene="P0401G10.6"
join(25336..25599,25803..26041,26224..26428,26676..26744,
26826..27012,27717..27883,28412..29002)
/gene="P0401G10.6"
/note="contains ESTs"
C73071(E2861), D46758(S11633), A0032617(S11633)
similar to Arabidopsis thaliana T22A6.120
unknown protein"
/codon_start=1
/protein_id="BAB88998.1"
/db_xref="GI:20146216"
/translation="MAKRAKLVQSEASIRISGLVDIANDIRLAKCKQSGPDLII
ELDKQVODIVLGNLIVTVGSKSIKDKENRFPNDVLSFQMSQREKRELSLCK
IPSGFNAFEKTECMQKDAITKSLAFDWCITLVTA,LSKHHILKHOAVST
MEPALRAFKEKTEHLYVGVKQKQVITLKOQSSLOADVQRLKEMSPQRLD
ANGSDI,SIADSVAKDKVARREGRLEFVSNPLNYSSENEVLMPPKRGSGDKDI
SHSMLTNVQAEPRVIVKSTPIPTSLINGPGGFLNHLINLIRKPREELHOLE
FOLRQWAPVSDI,PLPQRRKSSASL,PNILGPKLYCTNMNITOLEDTYRPOVE
AEIKRYEPISGMSRFVCTAPVDSDDSI,TVGAHLLEVSHQFKILIRLRFSPVC
NATSVKPEWMSGPNLQKSGSLISTL,ISTHRTAALKPAPRPAVNVINSNVYGGPVP
PVOTKILRFVDTTEMRGPDDLPGVWVSGAKLIERGKISLRVXISLTVNSPDE
FSPDEER"
complement(join(29649..29810,30106..30253,30610..30692,
31728..31819,31966..32037,32133..32223,32519..32559,
32678..32899,32926..32974))
/gene="P0401G10.7"
complement(join(29649..29810,30106..30253,30610..30692,
31728..31819,31966..32037,32133..32223,32519..32559,
32678..32899,32926..32974))
/gene="P0401G10.7"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB88999.1"
/db_xref="GI:20146217"
/translation="MEGLVPLNPLFNNKRVGGFYHNLPAFCPLVLPNSGMCPLRP
DHRSPFTIRAAAASPLIPNRRGGGDERDAIIPNCSRRRSIMTGSYRCADFCR
AQAALIGSPADFLERAKESVHVLVPGMPGIAVAFKDFVEELYENLGGQAVITVA
ADPHKTHNRMSVDPCFRCKMNLITRYGVNLGVLPELILNKKMSAKAIGTIAAS
SLSKGVSPFSFISLQASVTRGIMRLILPSPSVYAVEVTCGLMLLSEPPMNF
SNKQDIALEFVDHMGPLAHLLEVNSFPIRKLILLIYRS"
join(33330..33566,33690..33862,33943..34015,34105..34239,
34309..34519,34843..34991,35126..35224,35874..35946,
36133..36265,36557..36651,36731..36839,37093..37252,
37390..37551,37771..37942,38211..38377)
/gene="P0401G10.8"
join(33330..33566,33690..33862,33943..34015,34105..34239,
34309..34519,34843..34991,35126..35224,35874..35946,
36133..36265,36557..36651,36731..36839,37093..37252,
37390..37551,37771..37942,38211..38377)
/gene="P0401G10.8"
/note="DNA mismatch repair protein"
/codon_start=1

Db 283 TTGAGGATTAGCCAGTATTTCTACCTATGGCTTCGAGTGAGGCTTTGGCCACATA 342
 Qy 472 ACTTATGTTGGCCATGTTACCGTGACACAGATTAACAGAAAGCCAAATGACGGCTACAG 531
 Db 343 AGCCATGCGCTCATGTTACTATTAACAGAAAGAGCTGATGAAAGTGATGACAGA 402
 Qy 532 GTTCTTACAGAGATGTTGTAATGGAAATGAGCCTTACGCTTGGCTGGGTAAGGA 591
 Db 403 GCAAGTACTACAGTGAAGAAACGCGCCCTTAACCAATGCTGCAATCAAGG 462
 Qy 592 ACTCAAGTATGTTGAAATATTTTCAACATGATGAGCCCGGCAAGAAATCAAGG 651
 Db 463 ACCCAATGATGAGTGAAGAGACCTTTTTCACATATGACCAAGAGAAAGTTTAA 522
 Qy 652 AACCTCAATGATGAGTGAAGAGACCTTTTTCACATATGAGTGAAGAGTTCATC 711
 Db 523 AATCCAGTGAAGAAATATGGAATTTTGAAGTTGTTGGAGATTTACAGTACAAAT 582
 Qy 712 ATCAAGTATGCTTCTCTGCGAGAAAGCATGAGCCAAATAGACAGATGTTCTAGTGA 771
 Db 583 GCAGGATTAAGTTTCTCACTTAAAGAAAGAGAGACAGTACGATGTTAGACACTA 642
 Qy 772 AGTACATCTCAAGTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 831
 Db 643 CCCAATGCTCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 702
 Qy 832 CTCATGAAGAAATGAAGTTTCAATGATGATGATGATGATGATGATGATGATGAT 891
 Db 703 CTGATGAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 753
 Qy 892 TACATGTCATGAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 951
 Db 754 TACATGTCATGAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 813
 Qy 952 AGGCTGTAGTACGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1011
 Db 814 CGTCTGTAGTACGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 873
 Qy 1012 CCTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1071
 Db 874 CCCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 933
 Qy 1072 GTTAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1131
 Db 934 GTTAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 993
 Qy 1132 ACAATGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1191
 Db 994 CGGCTGTAGTACGTTGATGATGATGATGATGATGATGATGATGATGATGAT 1053
 Qy 1192 ACTCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1202
 Db 1054 ACCCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1064

RESULT 8
 LOCUS AR131133 2484 bp DNA linear PAT 16-MAY-2001
 DEFINITION Sequence 4 from patent US 6191268.
 ACCESSION AR131133
 VERSION AR131133.1 GI:14119458
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 2484)
 AUTHORS Listay R.M., Bronner, C.Eric., Baker, S.M., Bollag, R.J. and
 TITLE Kolodner, R.D.
 JOURNAL Compositions and methods relating to DNA mismatch repair genes
 FEATURES Patent: US 6191268-A 4 20-FEB-2001;
 location/Qualifiers
 1..2484
 source /organism="unknown"

BASE COUNT 720 a 535 c 594 g 635 t
 ORIGIN
 Query Match 12.8%; Score 319.8; DB 6; Length 2484;
 Best Local Similarity 57.7%; Pred. No. 1e-54;
 Matches 595; Conservative 0; Mismatches 427; Indels 9; Gaps 1;
 Qy 172 ATCCGAGGTTGAGAGAGTGTGTGTAACCGCATGCGGCGGGAGGTATGACAGCG 231
 Db 43 ATTGCGGCGGTGAGAGAGAGTGTGTGTAACCGCATGCGGCGGGAGGTATGACAGCG 102
 Qy 232 CGCTGCGGCGGTGAGAGAGTGTGTGTAACCGCATGCGGCGGGAGGTATGACAGCG 291
 Db 103 CCAGTGAATGCTTCAAGAGATGATGATGATGATGATGATGATGATGATGATGAT 162
 Qy 292 TCCGTCGCGGTGAGAGAGTGTGTGTAACCGCATGCGGCGGGAGGTATGACAGCG 351
 Db 163 CAAATGATGTTTAAAGAGAGAGTGTGTGTAACCGCATGCGGCGGGAGGTATGACAGCG 222
 Qy 352 ATCAGGTTGAGAGATGTTGCAATATGTTGCAAGAGCATGCTCAATGATGATGCA 411
 Db 223 ATCAGGAAAGAAATGCTGATGATGATGATGATGATGATGATGATGATGATGATG 282
 Qy 412 TACGAGATGCTGACACATTAATGATGATGATGATGATGATGATGATGATGATGAT 471
 Db 283 TTTGAGGATTTTACGATGATGATGATGATGATGATGATGATGATGATGATGAT 342
 Qy 472 ACTTATGTTGGCCATGTTACCGTGACACAGATTAACAGAAAGCCAAATGACGGCTACAG 531
 Db 343 AGCCATGCGCTCATGTTACTATTAACAGAAAGAGCTGATGAAAGTGTGATACAGA 402
 Qy 532 GTTCTTACAGAGATGTTGTAATGGAAATGAGCCTTACGCTGGGTAAGGA 591
 Db 403 GCAAGTACTACAGTGAAGAAAGTGAAGCCCTCAATACATGCTGATGATCAAGG 462
 Qy 592 ACTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 651
 Db 463 ACCCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 522
 Qy 652 AACCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 711
 Db 523 AATCCAGTGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 582
 Qy 712 ATCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 771
 Db 583 GCAGGATTAAGTTTCTCACTTAAAGAAAGAGAGACAGTACGATGTTAGACACTA 642
 Qy 772 AGTACATCTCAAGTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 831
 Db 643 CCCAATGCTCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 702
 Qy 832 CTCATGAAGAAATGAAGTTTCAATGATGATGATGATGATGATGATGATGATGAT 891
 Db 703 CTGATGAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 753
 Qy 892 TACATGTCATGAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 951
 Db 754 TACATGTCATGAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 813
 Qy 952 AGGCTGTAGTACGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1011
 Db 814 CGTCTGTAGTACGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 873
 Qy 1012 CCTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1071
 Db 874 CCCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 933
 Qy 1072 GTTAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1131
 Db 934 GTTAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 993
 Qy 1132 ACAATGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1191

Db 994 GGGGTCAGCAGACATCGAGAGCAAGCTCCTGGGCTCCATTCCTCCAGATGTAATTC 1053
OY 1192 ACTCAGGCAATT 1202
Db 1054 ACCCAGACTTT 1064

RESULT 9
AX214172
LOCUS AX214172 2484 bp DNA linear PAT 06-SEP-2001
DEFINITION Sequence 10 from Patent WO0159092.
ACCESSION AX214172
VERSION AX214172.1 GI:15524260
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Nicolai, N.C., Sass, P.M., Grasso, L., Vogelstein, B. and
Kinzler, K.W.
TITLE Methods for generating hypermutable microbes
JOURNAL Patent: WO 0159092-A 10 16-AUG-2001;
The Johns Hopkins University (US)
FEATURES
source 1. 2484
Location/Qualifiers
BASE COUNT 720 a 535 c 594 g 635 t
ORIGIN

Query Match 12.8%; Score 319.8; DB 6; Length 2484;
Best Local Similarity 57.7%; Pred. No. 1e-54;
Matches 595; Conservative 0; Mismatches 427; Indels 9; Gaps 1;

OY 172 ATCCGAGGTTGGAGAGTGGTGTGAACCCGATCGGGCGGGGAGGTGATCCAGCGG 231
Db 43 ATTCGGCGCTGGAGAGACAGTGTGAACCCGATCGGGCGGGGAGGTGATCCAGCGG 102
OY 232 CCGTGTGGCGGCTGAAGAGAGTCAATCCAGAACCCCTGAGCGGCGCCCTCCAGCGGC 291
Db 103 CAGCTAATGCTATCAAGAGATGATGAGAACTGTTAGATCAAAATCCCAAGATAT 162
OY 292 TCCGTTGGGCTGAAGAGAGGCTGCTCAAGCTCATCGAGTCTCCAGTACGCCATGCG 351
Db 163 CAGAGGATGTTAAAGAGAGAGGAGGCTGAAGTGTGATTCAGATCCAGAACAGTCCAGCGG 222
OY 352 ATCAGGTTGAGGATTTGGCAATATTTGGCAAGAGGATCTACTCAAAAGTTATCTGCA 411
Db 223 ATCAGAGAAAGAGATCTGATGTATGTATGTAAGGTTCACTACTAATGTAAGTCAAGTCC 282
OY 412 TACGAGGATTCGACACCATTAATCGATGGGCTCAGAGGGGAGGCTTTGGCTAGTATG 471
Db 283 TTGAGGATTTAGCCAGTATTTCTTACTATGAGCTTTCGAGGAGGCTTTGGCCAGCAT 342
OY 472 ACTATGTTGGCCATGTTACGCTGACACGATACAGAGAGGCAATTTGACGCGTACAGG 531
Db 343 AGCCATGTGCTCATGTTACTATTTACACGAAAGACGATGAGAAAGTGTGCATACAGA 402
OY 532 GTTCTTACAGAGATGTTGTATGAGAGAAAGGCTTAAGCCTTGGCGTGGTGAAGA 591
Db 403 GCAAGTACTACGATGAGAAAGTGAAGAGGCTTCTTAAGCAATGCTGCAATCAAGG 462
OY 592 ACTCAAGTATGTTGAAGAAATATATTTTACACATGTRAGCCCGCAAGAAAGATTTGCG 651
Db 463 ACCCAGATACAGGTGAGAGACCTTTTTCACACATAGCCACAGAGAGAAAGAACTTTAAA 522
OY 652 AACCTCAATGATGATACCCCAAGATCGTACACTCATGAGTGGCTTCCAGTCCATCAG 711
Db 523 AATCCAGTGAAGATATGAGAAATTTTGGAGTGTGGAGAGATTCATGATACAGAT 582
OY 712 ATCAACGTTACCTCTCTTGGAGAAAGATGAGCCCAATGAGCAAGATGTTATAGTGA 771

Db 583 GCAGGATTAATTTCTGATGTAAGAAACAGAGAGAGACAGTACGATGTAAGACACTA 642
OY 772 AGTACATCTCAAGTTAGATGATGATGAGAGTGTCTATGAGGCGCTTCTGTGCTCGAT 831
Db 643 CCCAATGCCCTAACCGTGAGCAATATTCCTGATCTTTGAAATGCTGTAGTGAGAA 702
OY 832 CTCATATGAATTAAGGTTTCAATGAGGATGCTGAGATTCATCTTCATCAAGATGAGT 891
Db 703 CTGATAGAAATTTGATG-----GAGATTAAGCCCTGATCAAAATGATAGT 753
OY 892 TACATCTCAATGCAAAATTAATGAGCAAGAGATGATGATGATGATGATGATGAT 951
Db 754 TACATATCAATGCAAAATTAATGAGCAAGAGATGATGATGATGATGATGATGAT 813
OY 952 AGGCTGTAGACTGATGCTGCTTGAAGAGACTTGAATTTGTTGCTGCAACATG 1011
Db 814 GCTGTGAGATCACTCTCTTGAAGAAACCATGAGAAAGTGTATGAGCCTATTTG 873
OY 1012 CCTCAAGATCCAAACCTTTCATATGATGATGATGATGATGATGATGATGATGAT 1071
Db 874 CCAAAACACAGACACCATTCCTGATGATGATGATGATGATGATGATGATGATGAT 933
OY 1072 GTTAATATACACCCACAGAGAGAGTGTGCTTTGATCAAGAGCTTATTTGAA 1131
Db 934 GTTAATGTCACCCACAGAGATGATGATGATGATGATGATGATGATGATGATGAT 993
OY 1132 ACAAATGAATGCTATGAGAGAAACTGATGATGATGATGATGATGATGATGATGAT 1191
Db 994 CCGGTCAGCAGACATCGAGAGCAAGCTCCTGGGCTCCATTCCTCCAGATGTAATTC 1053

OY 1192 ACTCAGGCAATT 1202
Db 1054 ACCCAGACTTT 1064

RESULT 10
AX234588 2484 bp DNA linear PAT 11-SEP-2001
LOCUS AX234588
DEFINITION Sequence 10 from Patent WO0162945.
ACCESSION AX234588
VERSION AX234588.1 GI:15593571
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Nicolai, N.C., Sass, P.M., Grasso, L., Vogelstein, B. and
Kinzler, K.W.
TITLE Methods for generating hypermutable yeast
JOURNAL Patent: WO 0162945-A 10 30-AUG-2001;
The Johns Hopkins University (US); Nicolaides, Nicholas, C. (US);
Sass, Philip, M. (US); Grasso, Luigi (US); Vogelstein, Bert (US);
Kinzler, Kenneth (US)
FEATURES
source 1. 2484
Location/Qualifiers
BASE COUNT 720 a 535 c 594 g 635 t
ORIGIN

Query Match 12.8%; Score 319.8; DB 6; Length 2484;
Best Local Similarity 57.7%; Pred. No. 1e-54;
Matches 595; Conservative 0; Mismatches 427; Indels 9; Gaps 1;

OY 172 ATCCGAGGTTGGAGAGTGGTGTGAACCCGATCGGGCGGGGAGGTGATCCAGCGG 231
Db 43 ATTCGGCGCTGGAGAGACAGTGTGAACCCGATCGGGCGGGGAGGTGATCCAGCGG 102
OY 232 CCGTGTGGCGGCTGAAGAGAGTCAATCCAGAACCCCTGAGCGGCGCCCTCCAGCGGC 291
Db 103 CAGCTAATGCTATCAAGAGATGATGAGAACTGTTAGATCAAAATCCCAAGATAT 162

QY 292 TCCGTTGCGGTGAAGAGCGGTGCGCTCAAGCTATCATCAGCTCCAGTATGACGGCATGGC 351
 DB 163 CAAAGTATTGTTAAAGAGGAGCGCTGAAGTTATTCAGATCCAAAGCAATGACCGCGG 222
 QY 352 ATGAGGTTGAGATGTTGGCAATATTTGGCAAGGATCTACCTCAAGTATATCGCA 411
 DB 223 ATGAGGAAAGAAATCTGGATATTTGATGGAAGGTCCTACTAGTAATGACAGTCC 282
 QY 412 TACGAGATCTGCGAGACCAATTAATGATGGGTTGACGAGGAGGAGGCTTTGGCTAGTAT 471
 DB 283 TTTGAGGATTTAGCAGTATTTCTACCTATGGCTTCGAGAGTGGAGGCTTTGGCAGAGATA 342
 QY 472 ACTTATGTTGGCATGTTTACCGTGCAGACATTAACAGAGGCCAATTTGACGGCTACAGG 531
 DB 343 AGCCAGTGGCTGATGTTACTATTAACAAACCAAGAGCTATGGAAGTGTGATACAGA 402
 QY 532 GTTCTTACAGATGATGTTGATGAGAAATGAGCCCTAAGCCCTTGGCTGGCGTGAAGA 591
 DB 403 GCAAGTTACTACATGAGAAAGTGAAGCCCTCTTAACCATATGCTGGCAATCAAGGG 462
 QY 592 ACTCAAGTCATGTTGAAATCTATTTTACACATGTTAGCCGCAAGAAACATTTGACG 651
 DB 463 ACCAGATCAAGCGGTGAGGAGCTTTTTCACATAGCCAGAGAGAAAGCTTTTAAA 522
 QY 652 AACTCAATGATGATCAACCAAGATCGTACCTTATCATGCTGGTTGCAATGCATCAC 711
 DB 523 AATCCAAAGTGAAGAAATATGAGAAATTTTGGAAATGTTGGAGATTCAGTACATCAAT 582
 QY 712 ATCAAGTATCTTCTTCTGCAAAAGATGAGCCCAATGACCAATGTTTCATAGTCA 771
 DB 583 GCAGGATGATGTTCTCAGTTAAATAAAGAGAGAGACAGTACTGATGTTAGACACTA 642
 QY 772 AGTACATCTCAAGGTTAGATGATGATGAGAGAGTGTGCTGCTTGTGCTGAT 831
 DB 643 CCCAATGCCCAACCGTGGAGCAATATTCGCTCATCTTTGGAATGCTGTAGTCGAGAA 702
 QY 832 CTGATGAATAAAGTTTCATATGAGATGCTGACAGATTCATCTTCAAGATGATGT 891
 DB 703 CTGATGAATAAAGTTGAGATGCTGAGATTAACCTTACCTTCAAAATGATGT 753
 QY 892 TACATGCAATGCAATATTTGAGCAAGAAAGATTAATGATGATCTTTCATATAATGAT 951
 DB 754 TACATATTCATCAATCACTACTAGTGAAGAGAGGCACTTCTTACTCTTCAATCAACAT 813
 QY 952 AGGCTGTAGACTGCTGCTTGAAGAAAGATTAATGATGATGCTGCTGCTGCAACATTTG 1011
 DB 814 CGTCTGTTAGATCACTCTCTTGAAGAAAGCCATGAGAAAGATGATGAGCCTATTTG 873
 QY 1012 CTTCAAGATCAACACTTTCATATATGATGATGATGATGATGATGATGATGATGAT 1071
 DB 874 CCCAAAGAACACACACCCATCTGCTGATGATGATGATGATGATGATGATGATGAT 933
 QY 1072 GTTAATATACACCAACCAAGAAAGATGATGATGATGATGATGATGATGATGATGAT 1131
 DB 934 GTTAATATGACCCCAACCAAGATGATGATGATGATGATGATGATGATGATGATGAT 993
 QY 1132 ACATTAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1191
 DB 994 CGGTTGACAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1053
 QY 1192 ACTGAGGATTT 1202
 DB 1054 ACCGAGACTTT 1064
 RESULT 11
 HS007343
 LOCUS HS007343 2484 bp mRNA linear PRI 29-NOV-1995
 DEFINITION Human DNA mismatch repair protein homolog (hMLH1) mRNA, complete cds.
 ACCESSION U07343
 VERSION U07343.1 GI:463988
 KEYWORDS

SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 2484)
 AUTHORS Bronner, C.E., Baker, S.M., Morrison, P.T., Warren, G., Smith, L.G., Lescoe, M.K., Kane, M., Earlbino, C., Liford, J., Linblom, A., Tanner, R., Kane, M., Bojia, R.J., Godwin, A.R., Ward, D.C., Nordenskiold, M., Fishel, R., Kolodner, R.D., and Liskay, R.M.
 TITLE Mutation in the DNA mismatch repair gene homolog hMLH1 is associated with hereditary non-polyposis colon cancer
 JOURNAL Nature 368 (6468), 258-261 (1994)
 MEDLINE 94195398
 PUBMED 8145827
 REFERENCE 2 (bases 1 to 2484)
 AUTHORS Morrison, P.T.
 TITLE Direct Substitution
 JOURNAL Submitted (02-MAR-1994) Paul T. Morrison, Molecular Biology Core Facility, Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 002115, USA
 FEATURES
 source
 location/Qualifiers
 1..2484
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="3"
 /map="3p22-23"
 1..2484
 /gene="hMLH1"
 22..2292
 /note="homolog of S. cerevisiae PMS1 (Swiss-Prot Accession Number P14242), S. cerevisiae MLH1 (Genbank Accession Number U07187), E. coli MMR1 (Swiss-Prot Accession Number P23367), Salmonella typhimurium MMR1 (Swiss-Prot Accession Number P14161) and Streptococcus pneumoniae (Swiss-Prot Accession Number P14160)."
 /codon_start=1
 /product="hMLH1"
 /protein_id="AAC50285.1"
 /db_xref="GI:463988"
 /translation="MSFVAGVIRLDETVVNRVIRAGEVIVORPAMIKEMENCLDAS
 TSQIVLKEBGLKLIQIDNGVIRKEDLDIVERTTSLQSFEDLASISTYFGE
 ALASISHVAHVTITTTADGKCAVRSYSGKIKAPKPDAGNOQIVTEDYFVA
 TRKALKNPSEVEGKILEVGRYSVHNAISGSKKOGTEVADVRLPNASTVDNRS
 IFGNASRELIEGCEKTLAFKNGYISNAYSVKCIPLFLINRHLSTSLRAI
 EYVAAYLRKNTHPFLYSLIEISPOVDVAVHTKHEVHRLSTSLRAI
 LGSNSRMVETQTLPLAGRSGEMKSTSLSSSTSGSDVYVHNAVHTSREK
 LDFLQPLSKPSSQPAVTEKTISSGRHQDEMELEPAPEVAANKNLSLEED
 TTKGTSEMESEKRPSTSNPKRRHRESDVEMVDDSRKEVTACTPRRRIINTSVLS
 LOBEINEOGHEVREMLHNSFVGCNPQWALQHOTKLTILNTTSLSELEFYQILY
 DFANFGLRLSEBAPLEDLAMALDSPEGWTEDGPKGLAEYIVLEFKKRAMLAD
 YSELEIDSEGNLGLPLIDNVYPLIEGLPIFLRLATEVNMDEKCEPSISKEM
 FYIRKQIYSEETLSGQSEVSGSIPNSKMTVEHIVYKALSHILPRHFTEDGNI
 LQLANLPDLTKVERC"
 BASE COUNT 720 a 535 c 594 g 635 t
 ORIGIN
 Query Match 12.8%; Score 319.8; DB 9; Length 2484;
 Best Local Similarity 57.7%; Pred. No. 1e-54;
 Matches 595; Conservative 0; Mismatches 427; Indels 9; Gaps 1;
 QY 172 ATCCGAGAGTTGAGAGAGTGGTGAACCGATCGCGGGGGGAGAGTATCCAGGG 231
 DB 43 ATTGCGGGCGTGGAGAGAGAGTGTGACCGCATCCGCGGGGGAAGTATATCCAGCG 102
 QY 232 CCGTGTGCGGAGTGAAGAGCTCATGAGAAGACGCTGAGCGCTCCAGCGCTC 291
 DB 103 CCACCTAATGCTATCAAGAAGATGATGAGAACTGTTAGCAAAATACACAAGTAT 162
 QY 292 TCCGTTGCGGTGAAGAGAGCGTCAAGCTCATCCAGGCTCCGATGAGAGCGCATGGC 351
 DB 163 CAAATGATGTTAAAGAGGAGGCGCTGAAGTTGATGATGATCAAGAACAAATGAGCACC 222

352 ATCAGTTGGAGTATGGCAATATGTCGAAAGGACATACCTCAAGTATCTCA 411
 223 ATCAGTAAAGAGATCTGGATTTGTATGTAAAGTCTACTACTAGTAACTGACATCC 282
 412 TACGAGATCTGCAGACCAATAAATGATGAGGGGTCAGAGGGGCTTGGCTATGATG 471
 283 TTTCAGAGATTTTCCAGATTTCTACATATGCTTCGAGGTCAGGCTTGGCCAGCAT 342
 472 ACTATGTTGCCATGTTACCTGACACAGATACAGAAAGGCAATTCGAGGCTCAGG 531
 343 ACCCATGTGGCTATGTTACTATTTACAGAAACAGCTGATGAAAGTGTGATCAGAA 402
 532 GTTCTTACAGAGATGCTATGAGGAAATGAGGATGAGGCTTGGCTGCTGAAAGCA 591
 403 GCAGATGCTACGATGAGAAATGAGGAGGCTTCTTAACATATGCTGCTGATCAGAA 462
 592 ACTCAAGTATGCTGAGAAATGCTATTTTCAACATATGATGAGGCTGCTGCTGATG 651
 463 ACCCATGATGAGGCTGAGGAGGCTTCTTACATATGCTGCTGATGAGGAGGAAAG 522
 652 AACTCAATGATGATGAGGCTGAGGAGGCTGATGAGGCTGCTGATGAGGAGGAA 711
 523 ATTCAGAGTAAAGATGAGGAAATTTGAGAGTGTGAGGATGATGATGATGATGAT 582
 712 ATCAAGTATGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 771
 583 GCAGGATGATGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 642
 772 ATGATGATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 831
 643 CCCAATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 702
 832 CTGATGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 891
 703 CTGATGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 753
 892 TACATGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 951
 754 TACATGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 813
 952 AGGTTGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1011
 814 CGTGTGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 873
 1012 CCTCAAGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1071
 874 CCCAATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 933
 1072 GTTATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1131
 934 GTTATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 993
 1132 ACATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1191
 994 CGGTTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1053
 1192 ACTGAGGATG 1202
 1054 ACCCAGACTTT 1064

RESULT 12 2530 bp mRNA linear PRI 12-JUL-2001
 LOCUS BC006850
 DEFINITION Homo sapiens, mult. (E. coli) homolog 1 (colon cancer, nonpolyposis type 2), clone MGC:5172 IMAGE:345158, mRNA, complete cds.
 ACCESSION BC006850
 VERSION BC006850.1 GI:13905125
 KEYWORDS MGC.
 SOURCE Homo sapiens.
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 2530)
 Strausberg, R.
 Direct Submission
 Submitted (27-APR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgaps-remail.nih.gov
 Tissue Procurement: ATCC

COMMENT

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
 Sequencing Center
 Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: villalona@bcm.tmc.edu

Villalona, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAC Plate: 3 Row: 1 Column: 24
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA 91: 466461.
 Location/Qualifiers

FEATURES

source

1..2530
 /organism="Homo sapiens"
 /db_xref="locusid:4292"
 /db_xref="taxon:9606"
 /clone="MGC:5172 IMAGE:345158"
 /tissue_type="Placenta, choriochorion"
 /clone_id="NH_MGC_10"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 51..2321
 /codon_start=1
 /product="mult. (E. coli) homolog 1 (colon cancer, nonpolyposis type 2)"
 /protein_id="AA06850.1"
 /db_xref="gi:13905126"
 /translation="MSFVAGYIRLDELTVNRILAGVIRIOPANAIEKMEIENDLAKS
 TSIOVIVEGGLIKLIDODNGTGRKEDIVICRFTTSKLOSPEDASISGFRE
 ALASISHVAHTTTTKADKCAIRASYSQKAKPKPCAGNGOITVEDLFYNA
 TRRLAKPSEIEGKILIEVGRISVHNAGISFVKKGGIYADYRTIPNASTYDNR
 IFGNVSRLEIEGKILIEVGRISVHNAGISFVKKGGIYADYRTIPNASTYDNR
 ETVYAAVLPKTHPELILSLISPDNDVYVPRKHEVHLERQVHISLIRKAI
 LGSNSRMYFTOTLPGIAGSGEMVSTISSTSGSDKRYAHQVARTSRECK
 IDAFLOPLSKRSOPAIIVEDKIDISGARQODEMELPAPAVAAKNOISLED
 TTKGSESEKRGPTSNPRKREHDSVEVEDSREKEMTAACTPRRIINTSVLS
 LOEINIEGHVILKEMHNSFVSCVNPQALAHOKIKYLINTKISLEEFYOLILY
 DPAFNGVLRSEPAFLPDLALADSPESGTEEDGREGIARIVLPEKLEKMLAD
 YFSLIDEGNLIIGLPLIDNYVPLDELPLFLRLATEVWMDKCFSLSECKM
 FYSIKQYISESTLSGOOSEVPSIPMSMKVTEHIVYALNRSHLLPPKHFEDGNI
 LQALNLPDYKFERC"

CDS

BASE COUNT

742 a 543 c 601 g 644 t

ORIGIN

Query Match 12.8%; Score 319.8; DB 9; Length 2530;
 Best Local Similarity 57.7%; Pred. No. 1e-54;
 Matches 595; Conservative 0; Mismatches 427; Indels 9; Gaps 1;

172 ATCCGAGTGGAGGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 231
 72 ATTCGGCGGCTGAGCAGACAGAGTGAACCGATCCGCGCGGAGAGTATTCACAGCG 131
 232 CCCTGCTGGCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 291
 132 CCAGCTAATGCTATCAAGAGATGAGTGAACCTGTTGATGAGCAAAATCCACAAGTAT 191
 292 TCCGTTCCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 351

OY 279 GCGCTCCAGCGCTCCGTTGGGGTGAAGAGCGTGGCTCAAGCTCATCAGCTCTCCGA 338
 DB 129 ATCCAGCCCAATCCAGGTGACAGGTGAGCGGCGGTGGAGTGGCTTCAATTCAGGA 188
 OY 339 TGAGCGCCATGAGCATCAGGTTTGGAGATTTGGCAATTTTGGCGAAGGCAATCTACTC 398
 DB 189 CAATGGCAGCGGGATACAGACGGGAGAGATCTGGCCATCTGCGAGCGGTTCAACACCTC 248
 OY 399 AAGTATTCATGATACGAGATCTGACAGCCATAAATGATGGGTTGAGAGGAGCG 458
 DB 249 CAGCTGACCCGCTTGAGAGATTTTCCAGATACCAATTTGGCTTCAGAGGAGAGC 308
 OY 459 TTTGGCTATGATGATTAATGCTGAGTGTACCGTGAACAGATACAGAGGCAAT 518
 DB 309 CTGGGACCATTAATGATGAGCGGACCTTAAGCATCCAGACCAAGAGGCTTACAGAGA 368
 OY 519 GCAGCGCTACAGGTTTCTTACAGAGATGGTAAATGAGAAATGAGCCCTTGGCG 578
 DB 369 GTGCGGCTACAGGCGGACCTACGCGGATGGCAAGCTCCAGGTTCAAGCCGCTGTGC 428
 OY 579 TGGGTGAAAGAACTCAAGTATGTTGAAATCTATTTTCAACATGTTAGCCGCA 638
 DB 429 CGGAACCAAGGACAGCATTTTGCATAGAGATCTGTTCTACAACTGCCCAAGAG 488
 OY 639 GAAACATTTGCAGAACTCCAGATGATGATACCCAGATCTGATGATGATGATGAT 698
 DB 489 GCAGGAGCTGCTGCGCGCGGAGGATTCAGAGATTTGCGAGGTTCTGGTATGTA 548
 OY 699 TGCAGTCAATCAGATCAGATTTCTTCTTCTGCAAGAAAGCATGAGCCATAGAGCAGA 758
 DB 549 CGGTGTCACAAATCCAAAGGTGGGCTTCACTCTCCGCAAGGAGAGCGCCAGCTCC 608
 OY 759 TGTTCATGATGCAATGATCTTCAAGTTAATGATGATGATGATGATGATGATGATG 818
 DB 609 CCGCGGACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 668
 OY 819 TGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 878
 DB 669 AATTTCCAGAGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 719
 OY 879 CAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 938
 DB 720 CGAGGCGGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 779
 OY 939 TTTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 998
 DB 780 GTTCATACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 839
 OY 999 CTCTGCAACATTTGCTTCAACATTTCAACATTTCAACATTTCAACATTTCAACATTT 1058
 DB 840 TGGCACAATCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 899
 OY 1059 AGAAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1118
 DB 900 ACACAACTTTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 959
 OY 1119 GCGTATATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATGAAATG 1178
 DB 960 AGAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1019
 OY 1179 CAGGATTTTCCAACTCAGGATTTAACTATATCAGGATTTGCTCAGCTAACCTAACCAAAA 1238
 DB 1020 GAGGAGCTTTACAAACAGCTAAGGCTACAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1079
 OY 1239 GGAATAGGTTTCTGAGGCGGCTATGGGTCTGGAACAAATTTCTGAGAG 1298
 DB 1080 GCGTGAATTAAGCAAGCAATTTT-----ATCCAGAGAGATGTTGGCTGA 1123
 OY 1299 CCAATATGTTGAGAGAGATCCAGCAATCTCAGAGATTTGCAACACTTACTGGCAGG 1358,
 DB 1124 CTGACTCCAGAGCAAAAGCTGATTAAGTTTGGGCGCTCTGCTCAAAAGTACTCAG 1183

OY 1359 GC-AATCTTCAATCTTCAAAAGAAATTTGATCTGTATCTGTATGATGATGATGATGAT 1417
 DB 1184 GCGTATCTCTCAAGTCTCTCCAGAGAGGCTCTGCTGCGCAGAGAGAGATTTGCGGTTA 1243
 OY 1418 CAGGAGAAACCAAAAGATGCTGTATTTGCAACGCTCTCATGAGCTCTGTGGAAA 1477
 DB 1244 CAGCGGCAAAATAATCAGAGAACTGCTTGAACACGCTGTGATGATGATGATGATGATG 1303
 OY 1478 TAGATTTAGCTTCAATCTGAGGCTTTTGGACATTTGCAAGAACTGCAATATGTTGAC 1537
 DB 1304 TGGAGCGGAGTGAAGGCTTCAAGTGTGAGAGACGCTCAAAACTTGTCTATGTGGGCT 1363
 OY 1538 TTTGGATGAAGCTTTGCTTGTATACACAAATFACCCTTTATCTGTAAATGTCG 1597
 DB 1364 GTGTGATGAGAGGCGAGCCCTGTTCCAGCATGAAACGCGCTTAAATGTGACACAC 1423
 OY 1598 TAAATATGATGAAGTATATGATGATGATGATGATGATGATGATGATGATGATGATG 1657
 DB 1424 GGTGCTTATGAGAGAACTGTTCTATACGCGGATGATTTACAGATTTCAAAACTGCTCG 1483
 OY 1658 CTATTCAGCTGATGAACACAGCTTCCACTTCAAGAGTGTGATGATGATGATGATGATG 1717
 DB 1484 AATATCAATTAAGTCCGCAATGCTGCTCAAGGAGCTGCTATCTGCTGCGGAAAGCG 1543
 OY 1718 ATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1777
 DB 1544 AAGCGCGCGGTGAGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1603
 OY 1778 AGATCTTAAAGAAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1837
 DB 1604 ACATCTTCTGAAAGAGCGCGCCATCATGAGAGATTTTGGGCTGCGCATCTCGGAGG 1663
 OY 1838 ATGCAAAATGCAAGAGCTTCTGTTTGTATGAGACAGTACACCCCTGATATGAGAGCTC 1897
 DB 1664 ATGAATGCTGAATCTGCTGCGGAGTGTGCTGACACAGCATGACCTTGTGCGCCACC 1723
 OY 1898 TTCCAGATTTGCTTGGCTTTAGAAATGATGATGATGATGATGATGATGATGATGATG 1957
 DB 1724 TGGCTTATCTGCTGCTGCGCTTGGCCAGGAGTGAATGATGATGATGATGATGATGATG 1783
 OY 1958 TCAGAACAGT 1967
 DB 1784 TCGAGACAT 1793

RESULT 14
 BC021815 2548 bp mRNA linear ROD 07-AUG-2002
 LOCUS Mus musculus, clone MGC:19289 IMAGE:4020489, mRNA, complete cds.
 DEFINITION
 ACCESSION BC021815.1 GI:18255307
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 2548)
 REFERENCE
 Strausberg, R.
 Direct Submission
 Submitted (18-JAN-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 CONTACT
 Contact: MGC help desk
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

REMARK
 COMMENT

Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
 Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
 Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the J.M.A.G.E. Consortium/MLN at: <http://Image.lln.gov>
 Series: IRK Plate: 24 Row: b Column: 13
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis, Genomescan gene prediction, similarity but not identity
 to protein.

FEATURES

source

Location/Qualifiers
 1. 2548
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /map="CZECH 11"
 /clone="MGC:19289 IMAGE:4020489"
 /tissue_type="mammary tumor metastasized to lung."
 /mvtv_ltr/Mvtl model. Expression driven by an Mvtv-LTR
 enhancer."
 /clone_11b="NCI CGAP_Lu30"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6".
 39. 2321
 /codon_start=1
 /product="unknown (protein for MGC:19289)."
 /protein_id="AAH21815.1"
 /db_xref="GI:18255308"
 /translation="MAFYAGYIRLDEYVNRIRAGVYQIRANAIKEMICLDAKS
 TNIOYVKEGGLKLIQIDNNGIRKEDLDVCEFTYSKLTQFEDLASISTYFGE
 ALASISHVAHYITTKADPKCAVASYSDGLAPPRPCVGNQGLTLVDELFNRI
 TRKMLKNPSEYGRILEVYGRSINHSGISISVKGQETVSDVDTLNPATVDNIS
 IEGNVSRELLEVEGCEDETLAFKNGYSINAKSVKCKFLFIHRLVESALSKAI
 EYVYAVLPKMTNHPRLYSLTSLSPONVNVYVQVHEVHLHEESITLQRYOCHESKL
 LGSNSRMTFTOTLLPGLAGRSGEARPTTVGASSTYSGSDKTYATQVNRIDSQK
 IDAFLQVSLVSPQDPAPVPRGARTGSPERARREDEMLAPPAVARSNLE
 RESLMESTDITQAKAPTSPPSSRRRHEDSDVEENASGEMTAPVPRRIINTL
 SVLSQELISECHETLERLNLNHSFVGCVPQMLAHOQLYLLNTTKISEEFYQ
 ILIYDFANGVRLSEPAFLPDLMLALDSPESGTEDDGPEGLAEYIYELKKAE
 MLADFEVIDEENLIGLPLIDSVVPLRGLPIFIIRLAEVMMDEKKEFESLSK
 ECAMFYSTRKOYIIEESTLSGQSDMPSTSKPMKTYVEHIIYKAFRSHLPPKHYTE
 DGNVLQNLNPLIKVFERC"

CDS

BASE COUNT 709 a 599 c 651 g 589 t
 ORIGIN
 Query Match 12.0%; Score 300.6; DB 10; Length 2548;
 Best Local Similarity 56.5%; Pred. No. 8.2e-51;
 Matches 583; Conservative 0; Mismatches 439; Indels 9; Gaps 1;

172 ATCCGAGGTTGGAGAGTCGGTGTGTAACCGCATCGCGGGGAGGTGATCCAGCG 231
 11
 60 ATTGCGCGCTTGAGAGAGAGCGGTGTGAACCGCATAGCGGGGAGGTGATCCAGCG 119
 232 CGTGTGCGGCGGTGAGAGAGCTCATGAGAACGCTGAGCGTGGCGCTCCAGCGTC 291
 11
 120 CCGGCCAATGCTATCAAAAGATGATAGAAACTGTTAGATGCAAAATCTACAATATT 179
 292 TCCGTGGGTGAAGGAGCGGCTCAAGCTCATCCAGTCCCATGACGCCCATGCG 351
 11
 180 CAAAGTGTGTTAGAGAGAGGCGCTGAAGCTTAATTCATCCAAATGCGCACTGGA 239
 352 ATCAGGTTGAGAGATTGGCAATATTGTGCGAAAGCATCTACTCCCAAGTATCTGCA 411
 11
 240 ATCAGAGAGAGAGATCTGATATGTGTGTGAGAGAGTTCACTACGAGTAACTGCAACT 299
 412 TACGAGATCTCGACACCAATAATGATGGGGTTTCAGAGGGAGCGTTGGCTACTATG 471
 11
 300 TTGAGAGATTAGCCAGTATTCTACCTATGGCTTTCGTGTGAGCAATTTGGCAACATA 359
 472 ACTTATGTTGGCAGATGTTACCGTGACAAAGCATTAACAGCAATTCACAGCGCTACAG 531
 11
 360 ACCCATGTGGCCATGCTACTATTACMACCAAAACAGCTGATGGGAATGTGGTACAGA 419

532 GTTCTTACAGAGATGTTGTAATGGAAATGAGACCTTAAGCTTGGCCGTGGTGAAGA 591
 11
 420 GCAAGTTACTAGATGGAAGAGTGAAGGCCCTCTTAACCCCTGTGACAGCAACGAGG 479
 592 ACTCAAGTATGTTGTAATATCTATTTACACAGTGTAGCCCGCAAGAAACATTGCA 651
 11
 480 ACCCTATACGGTGTGAAGACCTTTTTCACATATATCAACAAAGGAGAAAGCTTAA 539
 652 AACTCCATGATGACATACCCCAAGATGTAGACTTCATGACGTGGTGGTGGTACATC 711
 540 AATCCAGTGAAGATGACGAAATATTTGGAAGTGTGGCAGGATTCATACACAT 599
 712 ATCAAGTTACCTTCTCTTGACAGAAAGATGAGCCCAATAGACAGATGTTCAATGCA 771
 600 TCAGCATTTAGTATCTCAGTTAAATAAACAAGGTGAGACAGTATCTGATGACAAAC 659
 772 AGTACATCTCAAGGTTAGATGCTATCAGAGAGTGTATGAGGCTCTGTCTGCTGAT 831
 660 CCCAATGCCCAACCGGTGCAACATTCGCTCCATCTTTGGAATGCGGTTACTGAGAA 719
 832 CTCATGAATAAAGGTTTCATATGAGATGCTGACATTCATCTTCAAGATGATGAT 891
 720 CTGATGAAGTTGGGTGT-----GAGGATTAACCTTACTTCAAAATGATGCG 770
 892 TACATCTCAATGCAATATATGTGGCAAGAGATTAACATGATGTTTTCATTAATGAT 951
 771 TATATATGCAATGCAACATATTCAGTGAAGAAAGTGCATTTCTTACTTCAACACAC 830
 952 AGGCTGTGACATGTAAGTCTTTGAAAAGCATTTGAAATTTGTACTCTGCACATTTG 1011
 831 CGCTGTGTGAATACACTGCTTTCAGAAAAGCCATTGAAACTATATGCACTATTTTG 890
 1012 CCGCAAGCATCCAAACCTTTCATATGATGATGTCATCATCTTCATACATGACAGAC 1071
 891 CCCAAACACACACACCATCTCTGACCTCATGTTGAAATACAGCCCTCAGAACGAG 950
 1072 GTTAATATACACCAACCAAGAAAGAGTTAGCTTTTGAATCAAGAGCTATTATTTGA 1131
 951 GTCATATGACACCCCAACCAAGCAAGATTCACCTTCTGATGAGAGCATTCGACAG 1010
 1132 ACATATGAATATCTTTTGAAGAAAACGTATGATTTATATCAACAGATATTTCCA 1191
 1011 CGTGTGACAGAGCAATTTGAGCAAGCTGCTGCGCTCCAAATTCCTCCAGATATTTC 1070
 1192 ACTCAGGCAAT 1202
 1071 ACCCAGACCTT 1081

RESULT 15
 AF250844 2283 bp mRNA linear ROD 19-APR-2000
 AF250844 Mus musculus Mult homolog 1 protein (MLH1) mRNA, complete cds.
 AF250844.1 GI:7595953
 SOURCE
 ORGANISM
 Mus musculus.
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE
 Kurnaran, M. and Rao, M.R.S.
 Cloning of the cDNA of the Mult homolog, MLH1 from mouse testis
 JOURNAL
 Unpublished
 2. (bases 1 to 2283)
 Kurnaran, M. and Rao, M.R.S.
 Direct Submission
 Title
 Institute of Science, CNR Rao Circle, Bangalore, Karnataka 560 012,
 India
 FEATURES
 source Location/Qualifiers
 1. 2283

gene
CDS
/organism="Mus musculus"
/db_xref="taxon:10090"
/tissue_type="testis"
1..2283
/gene="MLH1"
1..2283
/gene="MLH1"
/codon_start=1
/product="Mult. homolog 1 protein"
/protein_id="AF64514.1"
/db_xref="GI:7595954"
/translation="MAFVAVYTRLDDETVNRIAGEVLTORPANAIKEMTENDLAKS
TNQVVEGGLKLIQIDNGTIRKEDDIVERPTTSKLOTFEDLASISTGPEGE
ALASISHVAHYITTTKADKCAVRSASDGLQAPKPCAGOGTILVIEDLPTNI
TRRAKLNPEEEYGLILEVGRYSINSGISFVKKQGEVSDVRLPNATVDNIRS
IFGNAVRELIEVGEDCKTLAKMNGYISNANSYKCIJFLINHLVBSALIRAI
EYVVAALPKRTHPLTSLSEISPOANDVAVHPTKHEVHLHEESILORVOHIESL
LGSNSRMFTQTLBGLAPSGEPAAPPTGVASSTSGSGEAVYAVOMVRTSDOK
LDAFLQVSSLVPSQPODDPAPVARGAREGPERATREDDEMLALPAPAEASENLE
RESIMETSDAOKAAPTSPSSSKRRHRESDVEMBNAGKEMTACTYPRRIINT
SVLSIOEISERCHETLREILRNHSFVCVNPQWALQHOTKLYLNTYLSLELFO
ILYDFANFVLRSEBAPLDLAMLALDPSGWEEDDQKGLAYIVELKKAE
MLADYFSEVIDEENLIGLPLIDSYVPLEGPIFLIRLATEVNMDEKECESLSK
ECAMFYSIRROYILEESTLSGQSDMGSTSKPMKTVHEHIIYKAERSHLPKHFE
DGNVLQANLPDIYKPERC"
BASE COUNT 625 a 543 c 592 g 523 t
ORIGIN

Query Match 12.0%; Score 299; DB 10; Length 2283;
Best Local Similarity 56.5%; Pred. No. 1.7e-50;
Matches 582; Conservative 0; Mismatches 440; Indels 9; Gaps 1;

OY 172 ATCCGAGGTGGAGAGTGGTGGTGAACCGCATCGCGGGGAGAGTATCCAGGG 231
DB 22 ATTCGGCGCTCTGAGAGCGGTAGTGAACCGCATAGCGGGGAGGTCTTCAAGGG 81
OY 232 CGGTGGTGGGAGAGAGGTCTGATCGAGACAGCTCGACGTGGCGCTCCAGCGTC 291
DB 82 CCGGCCAATGCTATCAAGAGATGATAGAAACTTTAGATGCAAAATCTCAAAATAT 141
OY 292 TCCGTTGCGGTGAGAGACGCGCTCAACATCATCCAGGTCTCCGATGACGGCATGCG 351
DB 142 CAAGTGGTGTAGAGAGAGTGGCTGAAGCTAATTCAGATCCAGACATGACACTGGA 201
OY 352 ATCAGGTGAGAGATTTGGCAATATTTGCGAAAGGCAATACCTCAAGATTTATCTGCA 411
DB 202 ATCAGAGAGAGATCTGATATGTTGTTGAGAGGTCTACTACAGATTAACAGCAACT 261
OY 412 TAGAGAGATCTGAGACATTAATTCGATGGGTTTCAGAGGGAGGCTTGGCTATGATG 471
DB 262 TTTGAGAGATTTAGCCAGTATTTCTACCTATGCTTCTGTTGAGGCAATGGCAAGCATA 321
OY 472 ACTTANGTTGGCCATGTTTACCGTGAACAGATTAAGAGAGGCAATTTGACGGCTACAGG 531
DB 322 AGCCATGTGGCCATGCTATTAACACCAAAACAGCTGATGGAATGTGCTACAGAGA 381
OY 532 GTTTCCTACAGAGATGCTGAATGAGAAATGAGCCTTACGCTTGGCTCGGTTGAAGA 591
DB 382 GCAAGTACTCAGATGAGAAAGTGAAGCCCTCTAAACCCGTGACAGCAACGAGGC 441
OY 592 ACTCAAGTCAATGTTGAAATCTATTTTACACATGCTGAGCCCGCAAGAAACATTGCA 651
DB 442 ACCCTGATCAAGGTGAGAAAGCTTTTTCACATATATCAACAGGAGAAAGCTTTAAA 501
OY 652 AACTCCATGATGATACCCCAAGATGATGATGATGATGATGATGATGATGATGATGAT 711
DB 502 AATCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 561
OY 712 ATCAAGTCACTTCTTTCAGAAAGCATGAGCAATAGACCAATGTTTCAATAGTGA 771
DB 562 TCAGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 621
OY 772 AGTACATCTCAAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 831

DB 622 CCCAATGCCACAAACCGTGGACAAACATTCGCTCATCTTTGGAAATGCGGTATGTCGAGA 681
OY 832 CTCATGAAATATAAGGTTTCATATGAGATGCTGCAAGATTTCAATCTTCAAGATGATG 891
DB 682 CTGATGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 732
OY 692 TACATCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 951
DB 733 TATATGCAATGCAAACTATTCAGTGAAGAGTGCATTTTCTACTCTTCAATCAACGAC 792
OY 952 AGGCTGTGAGTCTGATGCTGTTGAAAGAGTATGAAATTTGTACTCTGCAACATTC 1011
DB 793 GCTGTGTAATACAGCTGCTGCTGAGAAAGACATTTGAACGTGATATGACAGATCTTG 852
OY 1012 CCTCAAGCATCAACCTTTTCAATATCATGATGATGATGATGATGATGATGATGATGAT 1071
DB 853 CCCAAACACACACACACCTTCTGATCTGATGATGATGATGATGATGATGATGATGAT 912
OY 1072 GTTATATACACCCACCAAGAAAGGTTAGCTTTTGAATCAAGAGGTATTTATGAA 1131
DB 913 GTCATGTACACCCACCAAGCAAGATTCATCTTCTGACAGAGAGATTTGACAG 972
OY 1132 ACAATAGAAATGCTATTTGAGAAATCTGATGATGATGATGATGATGATGATGATGAT 1191
DB 973 CGTGTGACAGACACATTTGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1032
OY 1192 ACTGAGCATTT 1202
DB 1033 ACCGAGACCTT 1043

Search completed: April 4, 2003, 22:13:39
Job time: 7744 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 4, 2003, 18:33:37 ; Search time 626 Seconds
(without alignments)
8997.201 Million cell updates/sec

Title: US-09-954-950-1

Perfect score: 2501
Sequence: 1 cggcagcgagatttcgagtc.....aaaaaaaaaaaaaaaaaaaaa 2501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq_101002:*

- 1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2501	100.0	2501	24	AAD36728
2	926.2	37.0	2381	24	AAD36729
3	319.8	12.8	2484	16	AAO90814
4	319.8	12.8	2484	22	AAO94015
5	319.8	12.8	2484	22	AAH75044
6	319.8	12.8	2484	22	AAH75047
7	319.8	12.8	2484	24	AAD39200
8	319.8	12.8	2484	24	ABK86091
9	319.8	12.8	2525	16	AAO97525

10	319.8	12.8	2543	22	AAH34941
11	315.2	12.6	1995	23	ABL17901
12	259.4	10.4	2515	23	AAS91395
13	249.8	10.0	4052	23	ABL17900
14	243.8	9.7	2577	16	AAO90941
15	212.8	8.5	3218	22	AAH75039
16	212.8	8.5	3218	22	AAH75039
17	138.8	5.5	3716	23	ABL04226
18	132.6	4.9	1902	23	AAS54342
19	113.6	4.5	1890	23	AAS53227
20	113.6	4.5	1830121	17	AAT42063
21	102	4.1	549	21	AAC95307
22	99	4.0	2019	21	ABN67709
23	97.8	3.9	375	24	ABN21989
24	97.8	3.9	14273	19	AAV52173
25	97.6	3.9	1950	23	AAS55932
26	94.4	3.8	1950	23	AAS55471
27	91	3.6	2869	23	ABL25906
28	88.6	3.5	48275	21	AAH81501
29	88.6	3.5	349980	21	AAF21610
30	88.6	3.5	837096	21	AAH81489
31	84.6	3.4	2355589	24	ABA90521
32	84.4	3.4	5695	20	AAH20636
33	84	3.4	2133	23	AAS53161
34	83.4	3.3	17087	20	AAH12984
35	81.4	3.3	624	18	AAV74542
36	81	3.2	624	23	AAS50300
37	79.8	3.2	2007	23	AAS51625
38	79.8	3.2	2010	23	AAS54599
39	76.6	3.1	111309	20	AAH20250
40	75.6	3.0	1909	24	ABK74245
41	73.6	2.9	3063	16	AAO97526
42	73.6	2.9	3063	22	AAH75042
43	73.6	2.9	3063	22	AAH75042
44	73.6	2.9	3063	22	AAH75042
45	73.6	2.9	3063	24	ABK86089

ALIGNMENTS

RESULT 1	AAAD36728	standard; cDNA; 2501 BP.
XX	AAAD36728;	
AC	21-AUG-2002	(first entry)
XX		
DT	Rice MLH1 cDNA.	
XX		
DE	Rice: MLH1: gene mutation; mismatch repair system; transposon tagging; tisuue-specific promoter; herbicidal safener; male sterility; detection; pollen formation; base pair mismatch; gene; ss.	
KW		
KW		
XX		
OS	Oryza sativa.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	121..2295
FT		/tag= "a
FT		/product= "Rice MLH1 protein"
PN	WO200224890-A2.	
PD	28-MAR-2002.	
XX		
XX	18-SEP-2001; 2001WO-US29088.	
PF		
PR	18-SEP-2000; 2000US-233124P.	
XX		
XX	(PION-) PIONEER HI-BRED INT INC.	
PA		
XX		
PI	Manajan PB;	

Human colon cancer
Drosophila melanog
DNA encoding novel
Drosophila melanog
Murine wild type M
Nucleotide sequenc
Yeast MLH1 cDNA.
Drosophila melanog
Pseudomonas aerugi
Haemophilus influe
Haemophilus influe
Cat flea head and
Streptococcus poly
Human OREX polynuc
Streptococcus pneu
Streptococcus pneu
Streptococcus pneu
Drosophila melanog
N. meningitidis pa
Neisseria meningit
N. meningitidis pa
Genomic sequence o
Polynucleotide seq
Enterococcus faeca
Enterococcus faeca
Staphylococcus aur
Staphylococcus aur
Staphylococcus aur
Staphylococcus aur
Borrelia burgdorfe
Bacillus licheniffo
Human DNA repair p
Nucleotide sequenc
Human PMS1 cDNA.
Human mismatch rep
Human cDNA encodin

XX WPI: 2002-416283/44.
DR P-PSDB; AAE22978.
XX
PT Novel rice M1H1 ortholog nucleic acid molecule for increasing
PT efficiency of targeted gene mutation or homologous recombination in a
PT plant and for generating plants with reversible male sterility
XX
PS Claim 1; Fig 1; 90pp; English.
XX
XX The invention relates to isolated rice M1H1 orthologue nucleic acids. The
XX nucleic acid is useful for increasing the efficiency of targeted gene
XX mutation or homologous recombination in a plant, by transforming a plant
XX with expression cassette comprising the nucleic acid linked to a chemical
XX inducible promoter, transforming the plant with nucleic acid comprising
XX a sequence having a desired mutation or a sequence to be homologously
XX recombined, where the transformation occurs in the presence of chemical
XX compound capable of inducing the promoter and the plant's cellular
XX mismatch repair system is inhibited and selecting the transformed plants
XX that contained the mutation or homologously recombined nucleotide
XX sequence. The plant cellular mismatch repair system is inhibited through
XX the use of transposon tagging of an M1H1 gene, sense- and antisense-
XX suppression of an M1H1 gene, antibody binding to an M1H1 polypeptide or
XX its variant, and targeted mutagenesis of specific amino acid residues
XX encoded by an M1H1 gene. The nucleic acid is also useful for producing
XX a reversible male sterility in a plant, by transforming a plant with an
XX expression cassette comprising a lexa DNA binding site embedded in a
XX tissue-specific promoter that drives expression in the plant operably
XX linked to the nucleic acid when expressed disrupts pollen formation or
XX function through inhibition of the plant's cellular mismatch repair
XX system, transforming the plant with a second expression cassette
XX comprising a nucleotide sequence encoding a lexa repressor protein.
XX operably linked to a chemically-inducible promoter that drives expression
XX in the plant, and exposing the plant to a compound capable of inducing
XX the chemical-inducible promoter, to induce expression of lexa repressor
XX protein. The tissue-specific promoter is an anther-specific promoter
XX and the chemical-inducible promoter is a herbicidal seifer. The
XX polypeptide encoded by the nucleic acid is useful for detecting,
XX locating, or removing a base pair mismatch (SNP). The present sequence
XX is rice M1H1 cDNA.
SQ Sequence 2501 BP; 754 A; 511 C; 594 G; 642 T; 0 other:
Query Match 100.0%; Score 2501; DB 24; Length 2501;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 361 GAGGATTGGCAATATTGTCGAAAGGCACTACTCAAAAGTTATCTGATCGAGAT 420
Qy 421 CTGACAGACATTAATATGATGGGTTTCAGAGGGAGGCTTTGGCTAGTACTATGTT 480
Db 421 CTGACAGACATTAATATGATGGGTTTCAGAGGGAGGCTTTGGCTAGTACTATGTT 480
Qy 481 GGCCATGTTACCGTGACACGATTAACGAGGCCAATTGACAGGCTTACAGGTTTCTTAC 540
Db 481 GGCCATGTTACCGTGACACGATTAACGAGGCCAATTGACAGGCTTACAGGTTTCTTAC 540
Qy 541 AGAATGCTGTAAAGGAAATGAGGAGCCCTAAGCCTTGGCTGGGTAAGGAACTCAAGTC 600
Db 541 AGAATGCTGTAAAGGAAATGAGGAGCCCTAAGCCTTGGCTGGGTAAGGAACTCAAGTC 600
Qy 601 ATGTTGAAATCTATTATTCAACATGATGAGCCCGCAAGAAACATTCGAACTCCAAAT 660
Db 601 ATGTTGAAATCTATTATTCAACATGATGAGCCCGCAAGAAACATTCGAACTCCAAAT 660
Qy 661 GATGACTACCCCAAGATGTAAGTACTTATCAGTCGCTTTCAGTCATCAGATCAGCTT 720
Db 661 GATGACTACCCCAAGATGTAAGTACTTATCAGTCGCTTTCAGTCATCAGATCAGCTT 720
Qy 721 ACCCTTCTTCCAGAAAGCAATGAGCCCAATGAGCAGATGTCATGATGCAATATCC 780
Db 721 ACCCTTCTTCCAGAAAGCAATGAGCAGATGTCATGATGCAATATCC 780
Qy 781 TCAAGTTAGATGCTATCAGAGATGCTATGAGGCTTCTGCTGCTGATCTCATAGAA 840
Db 781 TCAAGTTAGATGCTATCAGAGATGCTATGAGGCTTCTGCTGCTGATCTCATAGAA 840
Qy 841 ATTAAGTTTAAATGAGGATGCTGAGATGCTCAATCTTCAAGATGAGTATCATCTCA 900
Db 841 ATTAAGTTTAAATGAGGATGCTGAGATGCTCAATCTTCAAGATGAGTATCATCTCA 900
Qy 901 AATGCAATTAATGTCGCAAGAAAGATTAACAATGATCTTTTCAATATGATAGGCTTGT 960
Db 901 AATGCAATTAATGTCGCAAGAAAGATTAACAATGATCTTTTCAATATGATAGGCTTGT 960
Qy 961 GACTGACTGCTTTGAAAGAGCTATTAATTTGTTGTTACTGCAACATTTCCCTCAAGCA 1020
Db 961 GACTGACTGCTTTGAAAGAGCTATTAATTTGTTGTTACTGCAACATTTCCCTCAAGCA 1020
Qy 1021 TCCAAACCTTCAATATCATGCTTCCATATCATCTTCCATGAGAACGTTGATTAATA 1080
Db 1021 TCCAAACCTTCAATATCATGCTTCCATATCATCTTCCATGAGAACGTTGATTAATA 1080
Qy 1081 CACCCCAACCAAGAAAGAGTTAGCCTTTTGAATCAAGGCTATTTTGAACAATTAAGA 1140
Db 1081 CACCCCAACCAAGAAAGAGTTAGCCTTTTGAATCAAGGCTATTTTGAACAATTAAGA 1140
Qy 1141 AATGCTATTAAGGAAAGAACTGATGAATTTCTAATACACAGGATTAATTCCTCAAGCA 1200
Db 1141 AATGCTATTAAGGAAAGAACTGATGAATTTCTAATACACAGGATTAATTCCTCAAGCA 1200
Qy 1201 TTAACCTTATCAGGATGCTCAACCTTAACCCACAAAGATTAAGTTTCTGAGGCCAGT 1260
Db 1201 TTAACCTTATCAGGATGCTCAACCTTAACCCACAAAGATTAAGTTTCTGAGGCCAGT 1260
Qy 1261 ATGGTTTCTGAACAAATCTCAAAAATTCCTGTTGAGCCCAATGCTGAGAACAGATCCA 1320
Db 1261 ATGGTTTCTGAACAAATCTCAAAAATTCCTGTTGAGCCCAATGCTGAGAACAGATCCA 1320
Qy 1321 CGCAATCATCTGGAAGAAATGACACCTACACGAGGCAATCTTCAAACTTTGAAAG 1380
Db 1321 CGCAATCATCTGGAAGAAATGACACCTACACGAGGCAATCTTCAAACTTTGAAAG 1380
Qy 1381 AATTTGATCTTGTATCTGTAGAAATGTTGTAGATCAAGAGAAACCAAAAGATGCT 1440
Db 1381 AATTTGATCTTGTATCTGTAGAAATGTTGTAGATCAAGAGAAACCAAAAGATGCT 1440
Qy 1441 GGTGATTTGTCAAGCCGCTATGAGCTCTTGTGGAATTAAGATTCTTCACTCTGCTG 1500
Db 1441 GGTGATTTGTCAAGCCGCTATGAGCTCTTGTGGAATTAAGATTCTTCACTCTGCTG 1500

QY 1501 CTTTGGACATTTGTCAGAACTGACATATGTTGACTTGCCGATGAAAGCCTTGTCTTG 1560
 DB 1501 CTTTGGACATTTGTCAGAACTGACATATGTTGACTTGCCGATGAAAGCCTTGTCTTG 1560
 QY 1561 ATTCACACCAATACCCGCTTATACCTTGTAAATGTGTAAATTTAGTAAAGACTTATG 1620
 DB 1561 ATTCACACCAATACCCGCTTATACCTTGTAAATGTGTAAATTTAGTAAAGACTTATG 1620
 QY 1621 TACCAGCAAGCTTGTGCGCTTTGGAACTTCAATGCTTTCAGCTGCACTGAACCAAGCT 1680
 DB 1621 TACCAGCAAGCTTGTGCGCTTTGGAACTTCAATGCTTTCAGCTGCACTGAACCAAGCT 1680
 QY 1681 CCACTTCAGAGAGTTGCTGTGATGACACTGAAAGACGATGAATGATGATGAAAG 1740
 DB 1681 CCACTTCAGAGAGTTGCTGTGATGACACTGAAAGACGATGAATGATGATGAAAG 1740
 QY 1741 GATGATGAGAAATGAGATGAGATGAGAAAGTAAACACTGAGTACTTAAAGAAATGCTGAG 1800
 DB 1741 GATGATGAGAAATGAGATGAGATGAGAAAGTAAACACTGAGTACTTAAAGAAATGCTGAG 1800
 QY 1801 ATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1860
 DB 1801 ATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1860
 QY 1861 GTTGTACTGACAGCAGTACACCCCTGATGATGATGATGATGATGATGATGATGATGATGAT 1920
 DB 1861 GTTGTACTGACAGCAGTACACCCCTGATGATGATGATGATGATGATGATGATGATGATGAT 1920
 QY 1921 GGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980
 DB 1921 GGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980
 QY 1981 GGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2040
 DB 1981 GGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2040
 QY 2041 TTTATACAGAAATATAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2100
 DB 2041 TTTATACAGAAATATAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2100
 QY 2101 GATGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160
 DB 2101 GATGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160
 QY 2161 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220
 DB 2161 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220
 QY 2221 TCAATGCAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2280
 DB 2221 TCAATGCAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2280
 QY 2281 TTTGAAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2340
 DB 2281 TTTGAAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2340
 QY 2341 AGTGTGTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2400
 DB 2341 AGTGTGTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2400
 QY 2401 ACTGAAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2460
 DB 2401 ACTGAAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2460
 QY 2461 TATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2501
 DB 2461 TATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2501

RESULT 2
 AAD36729
 ID AAD36729 standard; DNA; 2381 BP.

XX
 AC AAD36729;
 XX
 DT 21-AUG-2002 (first entry)
 XX
 DE Arabidopsis thaliana MLH1 DNA.
 KM MLH1: gene mutation; mismatch repair system; transposon tagging;
 KM tissue-specific promoter; herbicidal safener; male sterility; detection;
 KW pollen formation; base pair mismatch; gene; ds.
 XX
 OS Arabidopsis thaliana.
 XX
 FH Key Location/Qualifiers
 FT CDS 3..2216
 FT /tag- a
 FT /product- "Arabidopsis thaliana MLH1 protein"
 XX
 PN W0200224890-A2.
 PD 28-MAR-2002.
 XX
 PF 18-SEP-2001; 2001MO-US29088.
 XX
 PR 18-SEP-2000; 2000US-233124P.
 XX
 PA (PION-) PIONEER HI-BRED INT INC.
 PI Mahajan PB;
 XX
 DR WPI; 2002-416283/44.
 DR P-PSDB; AAE22979.
 XX
 PT Novel rice MLH1 ortholog nucleic acid molecule for increasing
 PT efficiency of targeted gene mutation or homologous recombination in a
 XX plant and for generating plants with reversible male sterility
 PS Disclosure; Page 86-88; 90pp; English.
 XX
 CC The invention relates to isolated rice MLH1 orthologue nucleic acids. The
 CC nucleic acid is useful for increasing the efficiency of targeted gene
 CC mutation or homologous recombination in a plant, by transforming a plant
 CC with expression cassette comprising the nucleic acid linked to a chemical
 CC inducible promoter, transforming the plant with nucleic acid comprising
 CC a sequence having a desired mutation or a sequence to be homologously
 CC recombined, where the transformation occurs in the presence of chemical
 CC compound capable of inducing the promoter and the plant's cellular
 CC mismatch repair system is inhibited and selecting the transformed plants
 CC that contained the mutation or homologously recombined nucleotide
 CC sequence. The plant cellular mismatch repair system is inhibited through
 CC the use of transposon tagging of an MLH1 gene, sense- and antisense-
 CC suppression of an MLH1 gene, antibody binding to an MLH1 polypeptide or
 CC its variant, and targeted mutagenesis of specific amino acid residues
 CC encoded by an MLH1 gene. The nucleic acid is also useful for producing
 CC reversible male sterility in a plant, by transforming a plant with an
 CC expression cassette comprising a lexa DNA binding site embedded in a
 CC tissue-specific promoter that drives expression in the plant operably
 CC linked to the nucleic acid when expressed disrupts pollen formation or
 CC function through inhibition of the plant's cellular mismatch repair
 CC system, transforming the plant with a second expression cassette
 CC comprising a nucleotide sequence encoding a lexa repressor protein
 CC operably linked to a chemically-inducible promoter that drives expression
 CC in the plant, and exposing the plant to a compound capable of inducing
 CC the chemical-inducible promoter, to induce expression of lexa repressor
 CC protein. The tissue-specific promoter is an anther-specific promoter
 CC and the chemical-inducible promoter is a herbicidal safener. The
 CC polypeptide encoded by the nucleic acid is useful for detecting,
 CC locating, or removing a base pair mismatch (SNP). The present sequence
 CC is Arabidopsis thaliana MLH1 DNA.
 XX
 SO Sequence 2381 BP; 721 A; 450 C; 548 G; 662 T; 0 other;

Query Match 37.0%; Score 926.2; DB 24; Length 2381;

QY	352	ATCAGGTTTAGAGATTTGGCAATATTTGGCAAAAGGCAATACCTCAAAAGTTATCTCA	411
Db	223	ATCAGGAAAAGAAGATCTGGATATTTGTATGTGAAAGGTTCTACTACTAATGTAACATGCAGTCC	282
QY	412	TACGAGGATCTGAGACCAATMAAATCGATGGGGTTCCAGAGGGAGGAGCTTTGGCTGTATG	471
Db	283	TTTGAGGATTTAGCAGATATTTCTTACCTATATGGCTTCGAGGTGAGGCTTTGGCCAGCAT	342
QY	472	ACTTATGTGGCCATGTTACCGTGCACAGATTAACGAAAGGCCAATTTGCAGGCTACAGG	531
Db	343	AGCCATGTGGCTCATGTATCTATTACAGGAAACAGCTGATGGAAGAGTGTCCATACGA	402
QY	532	GTTTCTTACAGAGATGGTGTATGGAAGAAATGAGACCTTAACCTTTGGCTCGGTTGAAAGA	591
Db	403	GCAAGTTACTACAGATGGAAGAACTGMAAGCCCTCCCTAAACCATGTGTGTGGCAATCAAGG	462
QY	592	ACTCAAGTCATGTTGMAAATCTATTTTACACATGATGGTACCCCGAGMAAAACATTTGAG	651
Db	463	ACCCAGATTCACGGGGAGGAGACCTTTTATTCACATATGACCAGAGGAGGAAACCTTTAAA	522
QY	652	AATCCCATATGATACATACCCCAAGATTCGTAAGATTTATCATGTCGGTTGCATCCATCAC	711
Db	523	AATCCAAAGTAAGAATATGGAAGAAATTTGGAAAGTTGTGGAGAGTATTCATCAATACAT	582
QY	712	ATCAAGCTACTCTCTCTCTTCCAGAAAGCATGAGCCCAATAGACAGATTTTCATGTGCA	771
Db	583	GCAGGCAATTTAGTTTCTTCAGTTAAACAAAGAGAGACAGTACGTATGATTTAGGACACT	642
QY	772	AGTACATCCTCAAGGTTATGATGCTATCAGGAGTGTCTATGGGGCTTCTGTCTGTATG	831
Db	643	CCCAATCTCCCAACCGGAGACATATTCGCTCCATCTTTGAAATATGCTGTATGTGAGAA	702
QY	832	CTCATGAATTAAGTTTATATATGAGGATCTGCAAGATTCATCTTCAAGATGATGCT	891
Db	703	CTGATGAGAAATTTGATGT-----GAGGATMAAAACCTTAGCTTCAAAATGAAATGCT	753
QY	892	TACATCTCAATGCAAAATTTATGTGGCAAGAGATTAACATATTTCTTTCATTAATGAT	951
Db	754	TACATATTCATGAAATCTACTCAGTGAAGAAAGATGCATCTTCTTACTTTCATCAACCAT	813
QY	952	AGGCTTGTAGACTGTACTGCTTTGMAAAGAGCTATTGAATTTGTACTCTGCAACATTTG	1011
Db	814	CGTGTGTAGATCAACTTCTCTTGAGAAAGCCATATGAAACAGTGTATCAGACCTATTTGG	873
QY	1012	CCCTAAGCATCCAAACCTTCTCATATACATGTCATATCTTCATCCATCAACAAACGCTGAT	1071
Db	874	CCCAAAACACACCCCATCTCTGTACTCTAGTTTGAATATGTCCTCCAGAAATGTGAT	933
QY	1072	GTTATATTTACACCCCAACCAAGAAAGAGTTAGCCTTTTGAATCAAGAGGATTTATGTA	1131
Db	934	GTTATATGTGCAACCCCAACAAAGCATGAATTCATCTTCTCTCAGAGAGAGCATCTCTGGAG	993
QY	1132	ACAATTAAGAAATGCTATTGAGGAAAACTGATGAATTTCAATTAACACAGATATTTCAA	1191
Db	994	CGGGTGCAGACGACATCGAGAGACAGTCTCTGGCTCCAAATTCCTCCAGATATGACTTC	1051
QY	1192	ACTCAGGCAATT 1202	
Db	1054	ACCCAGACTTT 1064	
RESULT 4			
AAQ94015			
ID AAQ94015 standard; cDNA; 2484 BP.			
XX AAQ94015;			
XX 15-JAN-1996 (first entry)			
XX Human mismatch repair pathway gene. MTH1.			
XX Mismatch repair; MSH2; primer; identification; defect; alteration;			

XX	cancer; tumour; vaccine; ds.
OS	Homo sapiens.
XH	key Location/Qualifiers
FT	CDS 21..3243
FT	*tag= "a
FT	/product= human_mismatch_repair_protein_MHL
PN	MO9514085-AZ.
XX	
PD	26-MAY-1995.
PF	17-NOV-1994; 94WO-US13385.
XX	
PR	13-JUN-1994; 94US-0259310.
PR	17-NOV-1993; 93US-0154792.
PR	07-DEC-1993; 93US-0163449.
XX	(UYVE-) UNIV VERMONT & STATE AGRIC COLLEGE.
PA	(DAND) DANA FARBER CANCER INST.
PI	Fisher R, Kolodner RD, Keenan RAG;
XX	
DR	WPI: 1995-200377/26.
P	P-PDB; MAR76071.
XX	Determining alteration in human mismatch repair pathways - used in
PT	the diagnosis, prognosis and therapy of cancers and in screening
PS	assays
XX	
SS	Example 13; Page 237-239; 256pp; English.
CC	RAO94015 is the human mismatch repair pathway gene MHL. Defects or
CC	alterations in such a gene result in the accumulation of unstable
CC	repeated DNA sequences, a feature of a number of different cancers.
CC	The identification of a defect in the mismatch repair pathway can
CC	be diagnostic of a predisposition to cancer and prognostic for a
CC	particular mammalian cancer e.g colorectal, ovarian, endometrial
CC	(uterine), renal, bladder, skin, rectal and bowel. The nucleotide
CC	sequences and polypeptides of the hMSH2 gene may also be used for
CC	therapy and in vaccines.
SQ	Sequence 2484 BP; 720 A; 535 C; 594 G; 635 T; 0 other:
OY	Query Match 12.8%; Score 319.8; DB 16; Length 2484;
OY	Best Local Similarity 57.7%; Pred. No. 9.3e-66;
Matches	595; Conservative 0; Mismatches 427; Indels 9; Gaps
OY	172 ATCGGAGGTGGAGGAGTGGTGTTGAACCGCATCGCGGGGAGTGTATCCAGCGG 231
DB	
OY	43 ATTGGCGCGCTGGACGACAGACTGTGTAAACCGCATCGCGGGGAGTGTATCCAGCGG 102
OY	232 CCGTCGTGCGGGGTGAAGAGCTCATCGAGAACAACCTCCAGCGTGGCGGCTCCAGCGTC 291
DB	
OY	103 CCAGTAATGCTATCAACAAGATATTGAGAACGTTAATGCAAAAATCACACAAGATT 162
DB	
OY	292 TCCGTGGCGGGAAGAGAGGTGGGCTCAAGCTCATCCAGGCTGCCATGACAGCGCATATGC 351
DB	
OY	163 CAAGGATGTTTAAAGAGGAGGCGCTGAAGTGTATTCAGATCCAAAGCAATGGACCGGG 222
DB	
OY	352 ATCAGGTTTGAAGATTGGCAATAATTGTGCAAAAGGCATCTACTCAAAGTTATTCGCA 411
DB	
OY	223 ATCAGCAAGAAGATCTGATATTGTATGTCAAAAGTTCACTACTAGTAAACATCGCATCC 282
DB	
OY	412 TACGAGGATTCGACACATATAATTCATGTGGGTTCAGAGGGAGGCGCTTGGTAGTATG 471
DB	
OY	283 TTGGGAGATTAGCCAGTATTTCTTACCTATATGCTTTTCAGAGTGAAGCTTTGGCCAGACA 342
DB	
OY	472 ACTTATTTGGCCATGTACCCTGACAAAGATAACAGAAAGGCCAATTGCACGCGTACAGG 531
DB	
OY	343 ACCCATGTGGCTCATGTACTATTCACAAAGCAAAACAGTGTATGAAAGATGTGATACAGA 402
DB	

Accession	Gene	Strain	Position	Sequence
532	gittc	TACAGAGATG	591	TTATGAGATAGAGCATGAGCTTAAGCCTTACGCTGGGTTGAAGA
403	GCAAGT	TACACATGGA	462	AACTGGAAGCCCTCTTAACCATGTGCTGGCATCAAGG
592	ACCTAC	GTGCTTGA	651	AAATCTATATTTTACAAACATGATAGCCCGCAAGAAAACATTCGAG
463	ACCCAG	TATCGGTG	522	AGAGACCTTTTTCATCAACATAGCAGAGAGAAAAGCTTTAAAA
652	AACTCA	TGATGACT	711	ATCCCAAGATCGTACCCCAAGATCGTAGCTTCATCATCGGTTTGAGATCCATC
523	AATCCA	AGTGA	562	AAATATGGAATTTTGGAAATGTTGGCAGGATATTCAGTACACAT
712	ATCAAG	TCTACCTT	771	CTGCAGAAAGCATGAGGCAATTAAGCAAGATGTTTCATATGCA
583	GCAGGC	ATTAGTTT	642	CTTCATGATTAACACAGAGACAGACATAGCATGATGATAGGACACTA
772	AGTACAT	CTC	831	AGATGCTATACAGGTTAGTATGCTATACAGAGTCTATAGGAGCTCTGTCGTTCTGAT
643	CCCATAT	GCCTTCA	702	ACCGGAGCAATATTTCCCTCCATCTTGGAAATAGCTTAGTAGCGAA
832	CTCATAG	AAATTA	891	AAAGGTTTCAATGAGATGCTGCAGATTCATCATCTTCAGATGATGATGAT
703	CTGATAG	AAATTTG	753	ATTTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
892	TACATCT	CAATGCA	951	AAATATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
754	TACATAT	CCATGCA	813	AAATATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
952	AGCGTGT	AGACCT	1011	ACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
814	CGTGTG	TGATAT	873	CACTTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
1012	CTCTAAG	CATCA	1071	CAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
874	CCCAAAA	CACAC	933	CCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
1072	GTTTAT	TATAC	1131	CCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
934	GTTTAT	TGTCAC	993	CCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
1132	ACAAT	TAGAA	1191	ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
994	CGGTC	GACGAC	1053	ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
5192	ACTCAG	CAAT	1202	
1054	ACCCAG	CACTT	1064	

XX 11-FEB-2000; 2000US-0181929
PR
XX
PA (UYJO) UNIV JOHNS HOPKINS.

XX Nicolaides NC, Sassi PM, Grasso L, Vogelstein B, Kinzler KW,
PI

XX WPI; 2001-514664/56
DR
DR P-PSDB; AAB85854.

aa Making hypermutable bacteria for biocatalysis, bioremediation and drug
 pt discovery, involves introducing polynucleotide comprising dominant
 pt negative allele of mismatch repair gene under regulatory sequence
 pt control -

XX Example 1; Page 44-45; 68pp; English
PS

The invention provides a method for generating a hypermutable bacteria. The method involves introducing a polynucleotide having a dominant negative allele of a mismatch repair (MMR) gene under the control of an inducible transcription regulatory sequence, into a bacterium. The cell becomes inducibly hypermutable. The method is useful to create desirable output traits for commercial applications, using dominant negative alleles of mismatch repair proteins. The mismatch repair gene is a Muth, Muts, Muhl or Muhl homologue and can be selected from PMS2, MLH1, MSH3, PMSR or PMSR homologue. The hypermutable bacteria is useful for the production, biocatalysis, bioremediation and drug discovery. It is also useful in manufacturing industry for the generation of new biochemicals useful for detoxifying noxious chemicals from by-products of manufacturing processes or those used, as catalysts, for remediation of toxins present in the environment including polychlorobenzenes, heavy metals and other environmental hazards for which there is a need to remove them from the environment. The hypermutable bacteria is also useful for screening novel mutations in a gene or a set of genes that produce variant siblings that exhibit a new output trait not found in wild type cells. The bacteria are also useful for producing increased quantity or quality of protein or non-protein therapeutic molecule e.g. penicillin G, Erythromycin and Clavulanic acid, by biotransformation. Dominant negative alleles of the MMR gene are useful for producing higher quantities of recombinant polypeptides. The present sequence represents a human MLH1 cDNA.

Sequence 2484 BP; 720 A; 535 C; 594 G; 635 T; 0 other;

Query Match	12.8%	Score 319.8	DB 22	Length 2484
Similarity	57.7%	Pred No. 9.3e-66		
Best Local				
Matches	595; Conservative	0	Mismatches 427;	Indels 9; Gaps 1

Oy	172	ATCCGAGGTTGAGAGCATGCGTGGTGAACCCGACATCCGGGGGGGGGATGATCCAGCG	231
Db	43	ATTTCGGCGCGGTGGACACGACATGGTGTGAACCGCATCCGGGGGGGAAATTATCCAGCG	102
Oy	232	CCGTCGTCGGCGGTGAAGAGCTCATCTCAAGAAACAGCCTCGACGCTGGCGCCTCCAGCGTC	291
Db	103	CCAGCTAATGCTATCCAAAGACATGATGTAGAACTGTTTAGATGCAAAATCCACAAGATAT	162
Oy	292	TCCGTTCCGGTGAAGAGCGGTGGCCTCAAGCTCATCCAGGTCTCCGATGACGGCATGGC	351
Db	163	CAAGTGATGTTTAAAGAGGAGCGCTGAAGTTGATTCAGATCCAAACAAATGGACACCGGG	222
Oy	352	ATCAGGTTTAGAGATTGGCCATATTTGGCCGAAGACGACTACTOCTAAAGTTATGCA	411
Db	223	ATCCGAAAAAAMATCTGGATATTTGATGTGAAGCTTCCACTACTAGTAAACTGCATCC	282
Oy	412	TACAGGATGTGGAGACATTAATGATGAGGGGTTCAGAGGGGAGCGTTTGGCTAGTATG	471
Db	283	TTTAGGATTTAGCCAGTATTTCTACCTATGGCTTTCAGATGAGGGCTTGGCCAGCAT	342
Oy	472	ACTATGTTGGCCATGTTTACCGTGACACAGATPACAGAAAGGCCAATTGCCACGCTAACG	531
Db	343	AGGCATGTGGCTATGTTTACTATTACACGAAACACAGCTGATGGAAATGTGCATACAGA	402
Oy	532	GTTTCTTACAGAGATGCTATATGAGAAATGAGCCCTTAAGCCTTGGCGTGGCGTGAAGA	591

Db	403	GCAAGTTACTCAGATCGAAGAACTGAAAGGCCCTCTCTTAACCATGTGCTGGCAATCGAAGG	462
QY	592	ACTCAGACTCATGGTTTAAAATCTATTTTACAAACATGATGAGGCCCAAGAAAACATTCGAC	651
Db	463	ACCCAGATCAGGGGTGAGAGACCTTTTTCACATATAGCCACGAGAGAAAAGCTTTAAA	522
QY	652	AATCCCAATATGATACACCCCAAGATCGTAGCTTCATCACTCGGTTCGATGCATCAC	711
Db	523	AATCCAGTGAAGATATGGAATATTTTGGAAAGTTGTGGCAGATATTCAGTCAACAAAT	582
QY	712	ATCAAGCTTACCTTCTCTCTCGAAGAAAGCATGTGAGGCCATATAGAGCAATGTTCATATGCA	771
Db	583	GCAGGCAATTTAGTTTCTCAGTTAAATAAACAAGAGACAGATGACTGATGTTAGACACTA	642
QY	772	AGTACATCTCCACAGGTAGATGCTATCAGAGATGCTATAGGGGCTTCGTCTGCTGAT	831
Db	643	CCCAATGCTCTAACCCGAGACAAATATTCCTCCATCTTTGAAATAGCTGATAGTCGAGAA	702
QY	832	CTCATAGAAATTAAGTTTATATGAGATGCTCGCAGATTCAACTTCTAAGATGAGATGT	891
Db	703	CTGATAGAAATTTGATGT-----GAGGATAAAACCTTAGCTTAAATGAAATGAT	753
QY	892	TACATCTAAATGCAATTTATGTGGCAAGAAAGATTTCAATGATTTCTTTCAATAATCAT	951
Db	754	TACATATCATATGCAAACTACCTCAGTAGAATAAGATGATCTTCTTACTCTTATCAACCAT	813
QY	952	AGCGTTGAGACTGACTGCTTGTAAGAAAGAGCATTTGAAATTTTGACACTCTGCAACATTTG	1011
Db	814	CGCTGGTAGATCACTACTCTTGAGAAAAGCCATGAAACAGTATGATACGACCATTTGG	873
QY	1012	CCTCAGACATCCAAACCTTTTCATATACATGTCCATACATCTTCATCAGAACATGAT	1071
Db	874	CCCAAAAACACACACCCCTTCTCTTACCTCAGTTTGAATATCAGTCCCGACGAATGTGAT	933
QY	1072	GTTAATATACCCCAACCAAGAAAGAGTATAGCTTTTGATTCAGAGAGGTATTTTGA	1131
Db	934	GTTAATGTGACCCCAACAAACATGAAATGTACATCTCTCTGACAGAGAGATCTCTGGAG	993
QY	1132	ACAATTAAGAAATGCTATTGAGGAAAAAAGCTGATGAATTTTAATCAACACAGATATTCGA	1199
Db	994	CGGGTGCACACACACATCGAAGAGAGCTCTGGGCTCCAAATTCCTCCAGAGATGATCTTC	1055
QY	1192	ACTCAGGCAAT	1202
Db	1054	ACCCGACATTT	1064

RESULT

ID: AAD39200 standard; cDNA; 2484 BP

AC AAD39200;

DT 04-OCT-2002 (first entry)

Human mismatch repair protein, MLH1 cDNA.

KW Hypermutable; antibody-producing cell; dominant negative allele; MMR
KW mismatch repair gene; genetic diversity; biochemical activity; human,
KW gene; SS.

OS Homo sapiens

Key	Location/Qualifiers
33	3292

```

/*lag = d
/product= "Human MTH1 protein"

```

XX
XX
DN
W0300237967-A1

16-MAY-2002.

1999

is resistant to selected microbe by introducing polynucleotide
comprising dominant-negative allele of mismatch repair gene into
mammalian cell

Disclosure: Page 18; 68pp; English.

The invention relates to making (M1) a mammalian cell hypermutable
comprising introducing a polynucleotide with a dominant-negative allele
of a DNA mismatch repair gene (e.g. Post meiotic segregation increased 1
and 2 (PMS1 and PMS2), Muts homologue 2 (MSH2) and MTH1 homologue 1 and 2
(MTH1, 2) into the mammalian cell which is hypermutable. Also included is
a homogenous composition comprising a cultured, hypermutable, mammalian
cell with a dominant negative allele of mismatch repair gene.
(M1) is useful for making a mammalian cell hypermutable. The
composition is useful for obtaining a mammalian cell that is resistant
to a selected microbe. (M1) involves growing a culture of mammalian
cells with a dominant negative allele of a mismatch repair gene,
exposing the cells to the selected microbe, and selecting the mammalian
cell resistant to the selected microbe. The hypermutable cell is
resistant to gram-negative and gram-positive microbe, protozoan,
bacteria or fungi. The microbial resistance is selected by isolating
and testing conditioned medium from the hypermutable cell. The
composition is also useful for obtaining a cell comprising a mutation
in a gene encoding an antimicrobial activity. (M1) involves growing a
culture of mammalian cells with the gene encoding the antimicrobial
activity, and a dominant negative allele of a mismatch repair gene,
selecting a cell comprising a mutation. The cell is examined to
determine whether the gene comprises a mutation by analyzing a
nucleotide sequence of the gene or mRNA transcribed from the gene, a
protein encoded by gene or its phenotype. The composition
is useful for identifying antimicrobial agents, microbe-specific toxic
molecules, and for producing new phenotypes of the cell. (M1) is useful
for creating genetically altered antimicrobial molecules, and also for
creating cell lines that manufacture antimicrobial agents for use in
large scale production of antimicrobial agents for clinical studies.
(M1) is also useful in cell lines that express known antimicrobial
agents to enhance the biochemical activity of the antimicrobial agent.
The present sequence encodes human MLH1.

Sequence 2484 BP; 720 A; 535 C; 594 G; 635 T; 0 other;

Query Match 12.8%; Score 319.8; DB 24; Length 2484;
Best Local Similarity 57.7%; Pred. No. 9,3e-66;
Matches 595; Conservative 0; Mismatches 427; Indels 9; Gaps 1;

172 ATCCGAGGTTGGAGAGCGGTGTGTAACCCATCGGGGGGGAGGTATCCAGCG 231
43 ATTCGGCGGTGGAGAGCGGTGTGTAACCCATCGGGGGGGAGGTATCCAGCG 102
232 CGGTCTGGCGGTGGAGAGCGGTGTGTAACCCATCGGGGGGGAGGTATCCAGCG 291
103 CCAGCTATGCTATCAAGAGATGATGAACTGTTAGATGCAAAATCCCAAGTATT 162
292 TCCGTTCCGTTGAGAGAGCGGTGTGTAACCCATCGGGGGGGAGGTATCCAGCG 351
163 CAGTATGTTTAAAGAGGAGCGGTGTGTAACCCATCGGGGGGGAGGTATCCAGCG 222
352 ATCAGGTTTATGAGTGTGTAACCCATCGGGGGGGAGGTATCCAGCG 411
223 ATCAGGTTTATGAGTGTGTAACCCATCGGGGGGGAGGTATCCAGCG 282
412 TACGAGGATGTCAGAGCATTAATGAGTGTGTAACCCATCGGGGGGGAGGTATCCAGCG 471
283 TTTAGATTTAGGCAATGTTTACCTATGCGTTTGAAGAGGCTTTGGGCAACANA 342
472 ACTTATGTTGCGCATTTACCGTGAAGAGTGAAGAGGCAATTTGACGGCTACAG 531
343 AGCCATGTGCTCATGTACTATCAACGAAAGACGTATGGAAGAGTGTGATACAGA 402
532 GTTCTTACAGAGATGTGTAAAGAGATGAGCTTACGCTTGGGTTGGAAGAG 591
403 GCAAGTTACAGATGAGTGAAGAGGCGGCTCTTAAACCATGTGCTGGCAATCAAGG 462

592 ACTCAAGTCATGCTTGAAGAAATCTATTTTACAAACAGTACCGCCGAGAAACATTCAG 651
463 ACCGAGATCAGCGTGGAGAGACCTTTTATACATAGCCAGAGAGAGAAAGCTTTAAA 522
652 AACTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 711
523 AATCCAGAGGAG 582
712 ATCAACGTTACCTTCTCTTCTGCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 771
583 GCAGGATTAAGTTTCTGCACTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 642
772 AGTACATCTCCAGAGTGTAGATGATGATGATGATGATGATGATGATGATGATGAT 831
643 CCCAATGCCCAACCGTGGAGATATTCCTCATCTTGGAAAGCTGTATGAGAGA 702
832 CTCATGAAATTAAGTTTTCATATGATGATGATGATGATGATGATGATGATGATGAT 891
703 CTGATAGAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 753
892 TACATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 951
754 TACATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 813
952 AGGCTTGTAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1011
814 CGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 873
1012 CTTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1071
874 CCCAATGCCCAACCGTGGAGATATTCCTCATCTTGGAAAGCTGTATGAGAGA 933
1072 GTTATATACACCCAG 1131
934 GTTATATGACCCAG 993
1132 ACAAATGAG 1191
994 CGGTTGACAG 1053
1192 ACTCAGGATTT 1202
1054 ACCCAAGCTTT 1064

RESULT 9
AA097525 standard; cDNA; 2525 bp.
ID AA097525;
AA097525;
25-MAR-1996 (first entry)
Human DNA repair protein hMLH1 coding sequence.
DNA repair protein: hMLH1; hMLH2; hMLH3; therapy; cancer; vectors;
DNA synthesis; diagnosis; disease; mutL4; ds.
Homo sapiens.
Key location/Qualifiers
CDS 42..2311
/*tag= a
/*product= DNA repair protein.
FT 318..320
FT misc_feature
/*tag= b
/*transl_except= AGT encodes Glycine.
FT 581..583
FT misc_feature
/*tag= c
/*transl_except= GCC encodes Glycine.
FT 911..913
FT misc_feature
/*tag= d

03-SEP-2001 (first entry)
Human colon cancer antigen encoding cDNA SEQ ID NO:2023.
Human colon cancer; colon cancer antigen; diagnosis; detection;
colorectal carcinoma; chromosome 3; ss.
Homo sapiens.
WO200122920-A2.
05-APR-2001.
28-SEP-2000; 2000WO-US26524.
29-SEP-1999; 99US-0157137.
03-NOV-1999; 99US-0163280.
(HUMA-) HUMAN GENOME SCI INC.
Ruben SM, Barash SC, Birse CE, Rosen CA;
WPI; 2001-235357/24.
P-PSDB; AAG73536.
Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
useful for preventing, diagnosing and/or treating colorectal cancers -
Claim 1; Page 3519; 9803pp; English.
AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
cancer-associated nucleic acid molecules (N) and proteins (P), where
the proteins are collectively known as colon cancer antigens. The colon
cancer antigens have cytostatic activity and can be used in gene
therapy and vaccine production. N and P may be used in the prevention,
diagnosis and treatment of diseases associated with inappropriate P
expression. For example, N and P may be used to treat disorders
associated with decreased expression by rectifying mutations or deletions
in a patient's genome that affect the activity of P by expressing of P.
Inactive proteins or to supplement the patient's own production of P.
Additionally, N may be used to produce the colon cancer-associated ps,
by inserting the nucleic acids into a host cell and culturing the cell
to express the proteins. N and P can be used in the prevention, diagnosis
and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
and AAG77789 represent sequences used in the exemplification of the
present invention.
N.B. Pages 666 to 682 and page 7053 of the sequence listing were
missing at time of publication, meaning no sequences are present for
SEQ ID NO:1027 to 1052, 7921 and 7922.
Sequence 2543 BP; 748 A; 548 C; 605 G; 638 T; 4 other;
Query Match 12.88; Score 319.8; DB 22; Length 2543;
Best Local Similarity 57.78; Pred. No. 9.4e-66;
Matches 595; Conservative 0; Mismatches 427; Indels 9; Gaps 1;
172 ATCCGAGGTTGAGAGAGTCGCTGTAACCGATCGCGGCGGAGGTGATCCACGG 231
81 ATTCCGCGCGCTGAGACAGACAGTGTGAAACCGATCGCGGCGGAGGTGATCCACGG 140
232 CCGTGTGCGGCGGTGAGAGAGTCATCGAAGACAGCTCGAGCGTGGCGCTCCAGCGTC 291
141 CCAGCTAATGCTATCAAGAGATGATGAACTCTTTAGATGCAAAATTCACAAATATT 200
292 TCCGTTGGGTTAAGAGAGCGTGGCTCAAGCTATCGAGTCTCGATGACGGCATGGC 351
201 CAAGGATGTTGTAAGAGAGGAGCGCTGAGTTGATTCAGATTCAGAAAGAGGACGGG 260
352 ATCAGGTTGAGGATTTGCAATATTGCGGAAGAGCATACTACTCAAAAGTTATCTGCA 411
261 ATCAGGAAGAAGATCTGATATTGATGTAAGAGTTCACTACTAGTAAACTGCAGTCC 320
412 TACGAGGATCTGACACCAATAAATGCAATGGGTTCAAGAGGAGGAGCTTTGGCTAGTAG 471

321 TTGAGGATTTAGCCAGTATTTCTACCTATAGCTTTGCGAGGTGAGCGCTTGGCCAGCAT 380
472 ACTTATGTTGGCCAGTTATACCGTGACACAGCATTAACAGAGCCATTTGACGGCTACAGG 531
381 AGCATTGTGGCTATGTTACTATTTACACAGAAAGAGCTGATGAAAGTGTGATACAG 440
532 GTTCTTACAGAGATGTTGTAATGGAATGAGACCTTAAGCTTGGCTGGCTGAAAGA 591
441 GCAAGTTACTACAGATGAAAGAACTGAAAGCCCTCTTAACCATGTTGCTGCAATCAAGG 500
592 ACTCAGTATGTTGTAATTTATTTTACACATGTTAGCCCGGCAAGAAACATTGAG 651
501 ACCAGATTCAGGTGAGAGACCTTTTACACATGATCCAGCAGAGGAGAAACCTTTAAA 560
652 AACTCCATATATACATACCCAGATCGTAGATCTATCATGCTTGGCTTGCATCAGAC 711
561 AATCCAGTGAAGATTTGGAATTTTGAAGTTGTTGAGATTTGAGATTTACATGATCAAT 620
712 ATCAAGTATCTTCTCTTCAGAAAGCATGAGCCATGAGCAGATGTTCTATAGTGA 771
621 GCAGCATTTAGTTCTCAGTTAAACAAAGAGAGACAGTATGATGTTAGACACTA 680
772 AGTACATCTTCAAGTTAGTATGCTATGAGAGATGCTATGAGGCTTGTCTGCTGAT 831
681 CCCAATGCTTCACCCGTCGACAAATATGCTCCATCTTTGGAATGCTGTTAGTCAGAA 740
832 CTCATGAAATTAAGGTTTCATATGAGATGCTGACATTCATTCATCAAGATGATGAT 891
741 CTGATGAAATGATGATG-----GAGATTAACCTAGCCTCAAAATGAATGGT 791
892 TACATCTCAATTAAGTATGTTGCAAGAGATTTCAATGATCTTTCTTCAATTAATGAT 951
792 TACATCTCAATTAAGTATGTTGCAAGAGATTTCAATGATCTTTCTTCAATTAATGAT 851
952 AGCTGTGAGTGTGATGCTTGTGAAAGAGCTTGAATTTGTTGTTGTTGTTGTTGTTG 1011
852 GCTGTGAGTGTGATGCTTGTGAAAGAGCTTGAATTTGTTGTTGTTGTTGTTGTTG 911
1012 CCTCAGCATCCAAACCTTTCATATGATGCTTTCATATCTTCCATAGAACAGCTGAT 1071
912 CCCAAACACACACCCATCTCTGATCCTGATGATGATGATGATGATGATGATGATGAT 971
1072 GTTATATATACCCCAAGCAAGAGAGGTTGAGCTTGTGATCAAGAGCTATTTGAA 1131
972 GTTATATATACCCCAAGCAAGAGAGGTTGAGCTTGTGATCAAGAGCTATTTGAA 1031
1132 ACAATTAAGAAATGCTATTTGAGGAAAGTGAATTTGATTAACACAGATATTTCAA 1191
1032 CGGGTGCAGCAGCATGAGAGCAAGCTCCGCGCTCCAAATTCCTCCAGATGATCTTC 1091
1192 ACTCAGGCTTT 1202
1092 ACCCAGACTTT 1102
RESULT 11
ABL17901
ID ABL17901 standard; DNA; 1995 BP.
XX ABL17901;
XX 26-MAR-2002 (first entry).
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO:5176.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX Drosophila melanogaster.
XX WO200171042-A2.

PD 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PMD, Myers EM;
 PI WPI; 2001-656860/75.
 DR
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1, SEQ ID NO 5176; 21bp + sequence listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB161676-AB163051), expressed DNA
 CC sequences (AB161640-AB161675) and the encoded proteins
 CC (AB161737-AB162072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 1995 BP; 486 A; 534 C; 561 G; 414 T; 0 other;
 Query Match 12.6%; Score 315.2; DB 23; Length 1995;
 Best Local Similarity 49.8%; Pred. No. 1e-64;
 Matches 901; Conservative 0; Mismatches 883; Indels 26; Gaps 3;

QY 699 TGCAGTCATCAGTCACTTCTCTGAGAAAGCATGAGCAGCAATAGAGCAGA 758
 DB 555 CGCTGTCCAGCAATCAAGGGGTGGCTTCACTTCCGCAAGCAGGAGAGCCGCAACTGCG 614
 QY 759 TGTTCATATGCAAGTATCATCTCCAGGTAGATGCTATCAGAGGTCTATGGGCTTC 818
 DB 615 CCGCGGACACCGAGTGGCCAGTTCGCCATCCGAAACATCCGATCTTTACGGGGCAGC 674
 QY 819 TGTCTGTCGTGATCTCATAGAAATTAAGTTTTCATATGAGATGCTCAGATCAATCTT 878
 DB 675 AATTTCAGAGAGCTGTGTGAGTTTACCGCAGCGATGAGGT-----TTACAGAGTT 725
 QY 879 CAGATGATGATGTTTACATCTCAAAATATATGTCGCAAGAGATTAACATGATCTT 938
 DB 726 CGAGGCCGAGTGTCTGTATCTACAGTTATCTTCCGCAAAAAGTGTGATGTCGT 785
 QY 939 TTTCAATTAATGATAGGCTTGTAGCTGTACGTGCTTGAAGAGCTTTGAATTTGGTA 998
 DB 786 GTTCATTAACCCAGGAGCTGTGTGAGTCCACAGCTCTAGAGCTCCGTGAGCTGATTTA 845
 QY 999 CTCTGCAACATTCCTTCACATCCAAACCTTTCAATACATGTCATACATCTTCCATC 1058
 DB 846 TGGCAGATTAAGTCCCGCGTGGCCACCATCTTTCGTATATAGAGCTGAGCGCGCC 905
 QY 1059 AGAACAGTGTGATTAATATACACCCACCAAGAAAGAGGTAGCCTTTGAATCAGA 1118
 DB 906 ACAGAACTTGAGAGTGTATCTCATCCACCAACATGAGTGTGATCTCTACACAGA 965
 QY 1119 GCGATATATTAAGCAATTAAGAAATGCTATTTGAGAAAACTGATGATTTATATAC 1178
 DB 966 AGAGATTTGTATGATGATTAAGACAGAGTTGAAGCCCGAGCTCTAGCAGCAATGTCAC 1025
 QY 1179 CAGGATATTCACAACTTCAGCATTAATATATCAGGATTTGCTCAACCTTAACCAAAA 1238
 DB 1026 GAGGAGCTTTACAAACAGCTAAGCTACAGAGAGCCCGAGCTGATGAACCCAGTT 1085
 QY 1239 GATTAAGTTCGTGAGCCAGTATGCTGTGAAACAAATCTCAAAAATTTCTGTGAG 1298
 DB 1086 GCGTATTAAGAAACGCAACGCAATTT-----ATCCCAAGAGATGTTGCGTA 1129
 QY 1299 CCAATATGTGAGAACATTCACAGCATTCATCGAAGAAATGTGACACCTACAGGACAG 1358
 DB 1130 CTGACTCAGAGACCAAAAGCTGATTAATTTTGGGCTCTGCTCAAAAAGTACTCAG 1189
 QY 1359 GC-AATCTTCAATCTTGAAGAAATTTGATCTTGTATGTAAGAAATTTGTAAGAT 1417
 DB 1190 GCGATCTCCAGTCTCTCCAGAGAACGTCGTGTGTCGCAAGAGAGATTTTGGCTTA 1249
 QY 1418 CAAAGAGAAACCAAAAAGATGCTGTGATTTGTCAACCGCATGAGCTCTTGTGAAA 1477
 DB 1250 CAGCGGCAAAAATATCAAGAAATGTGCTTGAACAGCTGTAGACATGAGAAAGCGC 1309
 QY 1478 TAGATTTAGCTTCCATCTGCGCTTTTGGACATTTGCAAGAACTGACATATGTGGAC 1537
 DB 1310 TGGAGCGGAGTGTGAGCTTCAAGTGAAGACAGCTCAAAAATCTGTATGTGGCT 1369
 QY 1538 TTGCGCATGAGAGCTTTGCTTTGATTAACAACAATATCCCTTTATCTGTAATGTGG 1597
 DB 1370 GTGTGATGAGAGGCGAGCCCTGTTCACAGATGAAGAGCCCTTTATATGTGCAACAC 1429
 QY 1598 TAAATATTAAGAACTTATATGACAGCAAGCTTTGTGCGCTTTTGGAACTTCATG 1657
 DB 1430 GGTGCTTTAGTGAAGAGCTGTATATCAGAGGATGATTTACAGATTTCAAAAGCTGCG 1489
 QY 1658 CTATTCAGCTCAGTGAACCAAGCTTCAAGAGTGTGTGTATGCTGCTGCTGAAAGCG 1549
 DB 1490 AATATCAAAATTAAGTCCGATGCTTCAAGAGGCTGCTAAATCTTCTCTGAAAGCG 1549
 QY 1718 ATGAATGATGAGATGAGAAAGATGATGAGAACTGGAAGATTTGCAAGATTAACCTG 1777
 DB 1550 AAGCGCGGCTGTAGCCCGGAGAGAGCGGATTAAGCTGTAGCTGTGCGAGGCTGCTGCG 1609
 QY 1778 AGATTAATAAAGAAATGCTGAGATGATTAATGAGTACTTTCTTATTCACATGATCAAG 1837

OY 1361 AATCTCAATCTTGAAAGAAATTTGATCTGTATCTGTAAAGAAATCTGTAAAGTCA 1420
 DB 1803 TATCTCTCAGTTCCTCCAGAGAGCGTCTGCTGCCAGAGAGATTTTCCGCTTAC 1744
 OY 1421 GGAGAAACCAAAAGATCTGCTGATTTGTCAAGCCGTCATAGAGTCTCTTGTGAAATAG 1480
 DB 1743 CGGCAAAAATCAGAGAAATGCGCTTTGACAGCGTCTAGACTAGAGAAAGCCGCTGG 1684
 OY 1481 AATTCAGCTTCATCTGCGCTTTTGGACATTTGACAGAACTGCATATGTTGACCTTG 1540
 DB 1683 AGCGCAGTGCAGCTTACGTAGAGAGACAGCTCAAAAACCTGTATGATGGGCTGTG 1624
 OY 1541 CCGATGAAGCTTTCTTGTATACAAACAAATACCCGTTTACTTGTAAATGTGTAA 1600
 DB 1623 TGGATGAGAGCGGAGCGCTGTTCCAGATGAAGACCGCTTATATGTGCAACACAGCT 1564
 OY 1601 AATATAGTAAAGACTTATGTACCAAGCTTTGTGCGCTTTGGAACTTCATGCTA 1660
 DB 1563 CGTTTATGAGAGACTTCTTATCAGCGGATGATTTAGAGATTTCAAAACCTGCTGGAAA 1504
 OY 1661 TTTACAGCTGAGACAGCTCCATCTTCAGAGATTTGCTGTATGTCACACTGAAAGAGATG 1720
 DB 1503 TCACATATAGTCCGCGCATTCCTTCAGAGAGCTGTATCTTGTGCGTGGAAAGCGAG 1444
 OY 1721 AATGATGATGATGAAAGATGATGAGAACTGGAAGATGCAAGTAAACACTGAGA 1780
 DB 1443 CCGCGCGCTGAGACCGCCGAAAGAGCGGAGATTAAGCTGAGCTGCGGAGAGTGGGAGAA 1384
 OY 1781 TACTAAAGAAATGCTGATGATTAATGATGATCTTTCTTATCATATGATCAAGATG 1840
 DB 1383 TTTCTTGTAAAGAGCGCCATCATGAGAGATATTTCCGGCTGCGCATCTGAGAGATG 1324
 OY 1841 GCAATTTGCAAGACTTCTGTTGTATGAGACAGATGACCCCTGATGAGACCGCTTC 1900
 DB 1323 GAATGCTCAATCTCTGCCAGATTTGCTCAGACAGATGAGCTTGTGTGGCCACTGCG 1264
 OY 1901 CAGAAATTTGTTGGCTTTAGGAAATGATGTTACTTGGATGACGAAAGAGTCTTCA 1960
 DB 1263 CTGTTTACCTGCTCGCTTGGCCAGGAGATGATGAGTGGAGAGACAGCATGCTTGG 1204
 OY 1961 GAACAGT 1967
 DB 1203 AGACATT 1197
 RESULT 14
 AAQ0941
 ID AAQ0941 standard; DNA: 2577 BP.
 XX
 AC AAQ0941;
 XX
 DT 05-MAR-1996 (first entry)
 XX
 DE Murine wild type MLH1 gene, a MutL homologue.
 XX
 KM mMLH1; wild type; MutL homologue; cancer diagnosis; mismatch repair;
 XX
 OS tumour; susceptibility; mutation detection; ss.
 XX
 Mus musculus.
 XX
 FH Key Location/Qualifiers
 FT CDS 3..2188
 FT /tag- a
 FT /note- "The above CDS is putative and the START
 FT codon is absent.
 FT The reading frame has to be shifted 5' by
 FT one base at position 944 to decode to the
 FT amino acid."
 XX
 PN MO9516793-A1.
 XX
 PD 22-JUN-1995.

XX
 PF 16-DEC-1994; 94MO-US14746.
 XX
 PR 09-DEC-1994; 94US-0352902.
 PR 17-DEC-1993; 93US-0168877.
 PR 08-MAR-1994; 94US-0209521.
 XX
 PA (DAND) DANA FARBER CANCER INST INC.
 PA (UNOR-) UNIV OREGON HEALTH SCI.
 XX
 PI Baker SM, Bollag RJ, Bronner CE, Kolodner RD, Liskay RM;
 XX
 DR WPI: 1995-231583/30.
 DR E-PSDB; AAR75789.
 XX
 XX
 PT Determn. of a mutation in a mutL homologue or gene prod. in a tissue
 PT used to diagnose cancer susceptibility, and to identify and
 PT classify a DNA mismatch-repair-defective tumour
 PS Example 2; Fig 12; 168pp; English.
 XX
 CC AAQ0941 encodes AAR75789 the predicted mMLH1, a MutL homologue. A
 CC mutation in a hMLH1 or hPMS1 nucleic acid isolated from a subject,
 CC can be detected by comparing it with an analogous segment of the
 CC claimed wild type mMLH1. This method can be used to diagnose cancer
 CC susceptibility, or to identify and classify a DNA mismatch-repair
 CC defective tumour.
 XX
 S0 Sequence 2577 BP; 705 A; 611 C; 644 G; 617 T; 0 other;
 Query Match 9.7%; Score 243.8; DB 16; Length 2577;
 Best Local Similarity 54.8%; Pred. No. 1.1e-47;
 Matches 532; Conservative 0; Mismatches 427; Indels 12; Gaps 2;
 OY 222 CCGTCGCGCGGTGAAGAGCTCATGCAAGACGCTGAGCGCTGCGCCCTCCAGCGTC 291
 DB 3 CCGGCAATGCTATCAAGAGATGATTAAGAACTTTAGTGAATGCAAAATATCAAAATAT 62
 OY 292 TCCGTTGGCGGTGAAGAGCGGTGCGCTCAAGCTCATCGAGTCTCCAGTACGCGCATGCG 351
 DB 63 CAAGTGTGTGTTAAGAGAGTGGCTGAGAGCTTAATGATGATCAAGAAATGCGCACTGGA 122
 OY 352 ATCAGTTTGGAGATTTGGCAATATTTGTGCGAAAGGCACTACTCAAGTATATCTCA 411
 DB 123 ATCAGAGAGAGAGATCTGATATTTGTGTGTGAGAGCTTCACTACAGATGAAGTGAAGCT 182
 OY 412 TACGAGATCTGACAGACCATTAATGATGAGGTTCAGAGGGGAGGCTTGGCTAGTATG 471
 DB 183 TTTGAGGATTTAGCCAGATTTCTACATATGCTTTCGTGAGTACATTTGGCAAGCATAT 242
 OY 472 ACTTATGTTGGCCATGTTTACCTGTGACCAAGATTAACAGAAAGCCATTTGACAGGCTACAG 531
 DB 243 ACTCATGTGGCCCATGCTCATCTTACCAACCAAGAGTGAAGTGAAGTGTGCGTACAGA 302
 OY 532 GTTCTTACAGAGATGATGATGATGAGGATGAGGATGAGCTTAAGCTTGGCTGCGTGAAGA 591
 DB 303 GCAAGTTACTCAGATGAGAAAGTGCAGAGCCCTCTTAACCTGTGCGAGGAGCAACAGGGC 362
 OY 592 ACTCAAGTCATGTTGAAATCTATTTTACCAACATGATGAGCCCGCAGAGAAATGTCAG 651
 DB 363 ACCCTGATCAGGTGAGAGACCTTTTTCACATTAATACCAAGAGAGAGAGAGCTTTAAA 422
 OY 652 AACTCAATGATGATCAACCCCAAGATCCGATGATGATGATGATGATGATGATGATGAT 711
 DB 423 AATCCAAAGTGAAGAGTGAAGAAATTTTGAAGTGTGAGAGATTTCAATACACAT 482
 OY 712 ATCAAGTACCTTCTCTTGCAGAAAGATGAGAGCCATGAGCAAGATGATGATGATGATGAT 771
 DB 483 TCAGGATATGATATCAATTAATAAAGAGAGAGAGATGATGATGATGATGATGATGATGAT 542
 OY 772 AGTACATCTCAAGAGTGAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 831
 DB 543 CCCAATGCCCAACACGCTGAGCAACATTCCTCATCTTTGAAATGCGGTAGTGCAGAA 602

```

OY 832 CTGATAGAAATTAAGTTTATATAGAGATGCTGACATTCATCTTCAATGATGAT 891
DB 603 CTGATAGAAATGTTGTTT-----GAGGATTAACCCCTAGCTTCAAAATGATGCG 653
OY 892 TACATTCGAATGAATATATATGTTGGCAAGAAGATTCAATGATCTTTTCATTAATGAT 951
DB 654 TATATATGATGCAATGAATATATGTTGAGAGAGATGATTTTCTTCTTCAATCAACCC 713
OY 952 AGGCTGTATGATGCTATGCTTTGAAAAAGATATGATTTGATCTGCTGCAATG 1011
DB 714 GCTGCTGATGAATGATGCTTGTGAGAAAGCCATTTGAATGATATGACGATCTATTG 773
OY 1012 CTTAAGACATTCGAATCTTCAATATATATGCTTCAATATCTTCCATGAGAACAGCTGAT 1071
DB 774 CCAAAAACACACACCCCTTCTGATCTGATCTGATTTGAATGACCCCTGAGAAC--GTGAC 830
OY 1072 GTTAATATACACCCACCAAGAAAGAGTTAGCTTTGATTAACAAGCGCTATATGAA 1131
DB 831 GTCAATGATACACCCACCAAGAAAGATTTCTGATGAGAGAGACATTTCTGACAG 890
OY 1132 ACAATAGAAATGCTATTTAGGAAAAATGATGATTTCTATATACACAGGATTTCCAA 1191
DB 891 CGTGTGACGACGACATTTAGAGAACCTGCTGCTGCTCAATTCCTCCAGATGATTTTC 950
OY 1192 ACTGAGGCAAT 1202
DB 951 ACCGAGACCT 961

RESULT 15
AAH75039
ID AAH75039 standard; cDNA; 3218 BP.
AC AAH75039;
XX
DT 29-OCT-2001 (first entry)
DE Nucleotide sequence of yeast mismatch repair protein MLH1.
XX MLH1; mismatch repair gene; MMR gene; hypermutable yeast; ss.
XX Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 241..2550
FT /tag= a
FT /product= "MLH1"
XX
XX WO200162945-A1.
XX
XX 30-AUG-2001.
XX
XX 21-FEB-2001; 2001MO-US05447.
XX
XX 23-FEB-2000; 2000US-0184336.
XX
XX (UYJO ) UNIV JOHNS HOPKINS.
XX (NICO/) NICOLAIDES N C.
XX (SASS/) SASS P M.
XX (GRAS/) GRASSO L.
XX (VOGE/) VOGELSTEIN B.
XX (KINZ/) KINZLER K W.
XX
XX Nicolaides NC, Sass PM, Grasso L, Vogelstein B, Kinzler KW,
XX
XX WPI: 2001-522830/57.
XX
XX P-PSDB; AAG63951.
XX
XX Making hypermutable yeast that exhibit novel selected output traits for
XX commercial applications, comprises introducing polynucleotide
XX containing dominant negative allele of mismatch repair gene
XX

```

```

PS Disclosure; Page 35-36; 60pp; English.
XX
XX The present sequence encodes yeast MLH1. MLH1 is a mismatch repair
XX (MMR) gene. The specification describes a method for making a
XX hypermutable yeast, comprising introducing a polynucleotide containing
XX a dominant negative allele of a mismatch repair (MMR) gene, into a
XX yeast, whereby the cell becomes hypermutable. The method is useful
XX to create desirable output traits for commercial applications, using
XX dominant negative alleles of mismatch repair proteins. The hypermutable
XX yeast is useful for production, biocatalysis, bioremediation and drug
XX discovery. It is also useful in genetic screens for the direct selection
XX of variant subclones that exhibit new output traits. The hypermutable
XX yeast is also useful in the manufacturing industry for the generation
XX of new biochemicals, for detoxifying noxious chemicals from by-products
XX of manufacturing processes or those used as catalysts, for remediation
XX of toxins present in the environment including polychlorobenzenes, heavy
XX metals and other environmental hazards for which there is a need to
XX remove them from the environment. The yeast is also useful for producing
XX increased quantity or quality of protein or non-protein therapeutic
XX molecule e.g., penicillin G, Erythromycin and Clavulanic acid, by
XX biotransformation.
XX
XX Sequence 3218 BP; 1080 A; 549 C; 644 G; 945 T; 0 other;
XX
XX Query Match 8.58; Score 212.8; DB 22; Length 3218;
XX Best Local Similarity 52.0%; Pred. No. 2,9e-40;
XX Matches 528; Conservative 0; Mismatches 482; Indels 6; Gaps 2;
XX
OY 182 TGGAGAGATGCTGATGACCGCATCGCGGGGGAGGTGATCCAGCGCGCTGCTGCG 241
DB 263 TTGATGATCATGATGTTAAACAAATTTGCTGAGTGAATCATATATATTCCTTAATG 322
OY 242 CGGTGAAGAGATCGATGAGAACAGCCCTGAGCGCTGCGCTCCAGCGCTCCCTTGGCG 301
DB 323 CTCTCAAGAAATGATGAGAAATTCATGATGAGTGAATGCTAATGATTTATTTCTAG 382
OY 302 TGAAGGACGTTGCTCAAGCTCATCCAGGCTCTCCATGACGCCATGCGATCAGTTTG 361
DB 383 TCAAGGAGAGGAGATTAAGTACTTCAATATACAGATACGAGATCGAATTAATAAG 442
OY 362 AGATTTGGCAATTTGCGGAAGGCAATCTACTCAATGATTCGATGACGAGATC 421
DB 443 CAGACCTGCGCATTTATGAGCGATTCACGAGCTCAATTTCAAAATTCGAAGTT 502
OY 422 TGCAGACCATTAATTCATGGGTTTCAGAGGGAGGCTTTGGTATGATGATGTTG 481
DB 503 TGAGTCAGATTCAAAGCATGATGATTCGAGAGAGAGCTTTGACAGATTCACATGIGG 562
OY 482 GCCATGTTACCGTGACACGATTAACAGAGGCCAATTTGACAGGCTTACAGGTTCTTACA 541
DB 563 CAAAGATCACAGTAACGAAAGTTAAAGAAAGACAGATGTCATGAGATTTTCATATG 622
OY 542 GAGATGCTTAATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 601
DB 623 CAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 682
OY 602 TGGTTGAATATCTATTTTCAACATGTTAGCCCAAGAAACATTCAGAACTCATG 661
DB 683 TAGTTGAAGACCTTTTTCATATTTCTTCAATTAAGGCGCTTAGAGTCCCATATAG 742
OY 662 ATGACTACCCCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 721
DB 743 ATGATATCTTAAATATTAATTTGTCGCGGCGATACGCCATTCATCAAGAGCATTTG 802
OY 722 CCTCTCTGAGAAAGATGAGGACCAATAGAGATGTTCAATAGTCAATGATCATCT 781
DB 803 GCTTTCTGTAAGAAAGTGGAGACTCTAATTTCTTTATTCATGATTAACCTTCATATA 862
OY 782 CAAGTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 841
DB 863 CAGTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 922
OY 842 TAAAGTTTCAATAGAGATGCTGACATTCATCTTCAAGATGATGATGATGATGATGATG 901

```

Mon Apr 7 09:23:29 2003

us-09-954-950-1.rng

Db 923 TTCATATCAGCAAGTAGAATTAA---CTGGAAGCGTGTGATGAAAGGTGTGA 979
QY 902 ATGCAATTATGTGGCAAGAG---ATTACATGATCTTTGATTAATGATAGGCTTG 958
Db 980 ATTGAATTTCAATCCAAAAGTCCATTTCTATTAATTTTTCATTAATATAGACTAG 1039
QY 959 TAGACTGTACTGCTTGAAGAAGCTATTGAATTTGTGACTCTGCACATTGGCTCAG 1018
Db 1040 TGACATGTGATCTTAAAGAGAGCTTGACAGCGTTTACTCCAAATTAATCTGCCAAGG 1099
QY 1019 CATCAACCTTTCATATACATGTCATCTTCATCAGACAGCGTGTGAATA 1078
Db 1100 GCTTCAGACCTTTTATTATTGGAAITGTATAGATCCGGCGCTGTGATGTTAAG 1159
QY 1079 TACACCCAGCAAGAAAGAGGTAGCGCTTTTGATCAGAGCGTATTATGAAACATAA 1138
Db 1160 TTCACCCGACAAAGAGAGAGTTCCTTCAGCCAGATGAGATCATAGAGAAATCG 1219
QY 1139 GAAATGCTATTGAGAAAACATGATGAATCTAATACAAACAGATATTCCAAACT 1194
Db 1220 CCAATCAATTGACAGCGGAATTAATCTGCCATGATATCTTACCTTCAAGGCT 1275

Search completed: April 4, 2003, 20:04:56
Job time : 662 secs

DR N-PSDB: AAD36728.

XX Novel rice MLH1 ortholog nucleic acid molecule for increasing
 PT efficiency of targeted gene mutation or homologous recombination in a
 PT plant and for generating plants with reversible male sterility

PS Claim 7; Fig 2; 90pp; English.

CC The invention relates to isolated rice MLH1 orthologue nucleic acids. The
 CC nucleic acid is useful for increasing the efficiency of targeted gene
 CC mutation or homologous recombination in a plant, by transforming a plant
 CC with expression cassette comprising the nucleic acid linked to a chemical
 CC inducible promoter, transforming the plant with nucleic acid comprising
 CC a sequence having a desired mutation or a sequence to be homologously
 CC recombined, where the transformation occurs in the presence of chemical
 CC compound capable of inducing the promoter and the plant's cellular
 CC mismatch repair system is inhibited and selecting the transformed plants
 CC that contained the mutation or homologously recombined nucleotide
 CC sequence. The plant cellular mismatch repair system is inhibited through
 CC the use of transposon tagging of an MLH1 gene, sense- and antisense-
 CC suppression of an MLH1 gene, antibody binding to an MLH1 polypeptide or
 CC its variant, and targeted mutagenesis of specific amino acid residues
 CC encoded by an MLH1 gene. The nucleic acid is also useful for producing
 CC reversible male sterility in a plant, by transforming a plant with an
 CC expression cassette comprising a LexA DNA binding site embedded in a
 CC tissue-specific promoter that drives expression in the plant operably
 CC linked to the nucleic acid when expressed disrupts pollen formation or
 CC function through inhibition of the plant's cellular mismatch repair
 CC system, transforming the plant with a second expression cassette
 CC comprising a nucleotide sequence encoding a LexA repressor protein
 CC operably linked to a chemically-inducible promoter that drives expression
 CC in the plant, and exposing the plant to a compound capable of inducing
 CC the chemical-inducible promoter, to induce expression of LexA repressor
 CC protein. The tissue-specific promoter is an anther-specific promoter
 CC and the chemical-inducible promoter is a herbicidal safener. The
 CC polypeptide encoded by the nucleic acid is useful for detecting,
 CC locating, or removing a base pair mismatch (SNP). The present sequence
 CC is rice MLH1 protein.

S0 Sequence 724 AA;

Query Match 100.0%; Score 3709; DB 23; Length 724;
 Best Local Similarity 100.0%; Pred. No. 1.8e-297;
 Matches 724; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDEPSPGGGCGGCEPPTIRLEESVYVNRILAGEVIOFPSAAYKELIENSIDAGASSVSA 60
 DB 1 MDEPSPGGGCGGCEPPTIRLEESVYVNRILAGEVIOFPSAAYKELIENSIDAGASSVSA 60
 QY 61 VKDGLKLIQVSDGHRGIRFEDLALICERHTTSKLSAYEDLOTIKSMGRGELASMTYV 120
 DB 61 VKDGLKLIQVSDGHRGIRFEDLALICERHTTSKLSAYEDLOTIKSMGRGELASMTYV 120
 QY 121 GHVYTTTTEGOLHGRYSYRGVNEPEKPCAAYKGYVYENLFYNNVAKKTLQNSN 180
 DB 121 GHVYTTTTEGOLHGRYSYRGVNEPEKPCAAYKGYVYENLFYNNVAKKTLQNSN 180
 QY 181 DDPYKIVDFISRFVAHINVTFSCKRHGANRADVHASASTSRDLAIRSYGASVARDLIE 240
 DB 181 DDPYKIVDFISRFVAHINVTFSCKRHGANRADVHASASTSRDLAIRSYGASVARDLIE 240
 QY 241 IKVSTEDAADSIFFKMDGYISNANYAKKITMILFTNDRLVDTALRAIEVYSATLPOA 300
 DB 241 IKVSTEDAADSIFFKMDGYISNANYAKKITMILFTNDRLVDTALRAIEVYSATLPOA 300
 QY 301 SKPFIYMSIHLPSEHVDVNIHTKREVSILNDRILIEIRNAIEEKLNMSNTRTFQDA 360
 DB 301 SKPFIYMSIHLPSEHVDVNIHTKREVSILNDRILIEIRNAIEEKLNMSNTRTFQDA 360
 QY 361 LNLGSIAGANPOKDKVSEASMSGSKSOKIPYSQVWRDPRRSGRLRHVYWGSSNLEK 420
 DB 361 LNLGSIAGANPOKDKVSEASMSGSKSOKIPYSQVWRDPRRSGRLRHVYWGSSNLEK 420

QY 421 KEDLVSVNRVNRNRNCKADGLSSRHLELVEIDSSFFHGLLDIVKNTYGLADEAFAL 480
 DB 421 KEDLVSVNRVNRNRNCKADGLSSRHLELVEIDSSFFHGLLDIVKNTYGLADEAFAL 480
 QY 481 IOHNTRLVNVVNIKSKELMQALCRGNNAIQLSEPAIQLLWALKDDELMSDEK 540
 DB 481 IOHNTRLVNVVNIKSKELMQALCRGNNAIQLSEPAIQLLWALKDDELMSDEK 540
 QY 541 DDEKLEIAEVNTEILKEAEMINEVESIHIDQDKLTPVLVLOQYPPDMRLPEFVAL 600
 DB 541 DDEKLEIAEVNTEILKEAEMINEVESIHIDQDKLTPVLVLOQYPPDMRLPEFVAL 600
 QY 601 GNDVYWDDEKECFRTVASAVNFYALHPPIIPNSGNGIHLYKKNRDSMADEHAENDLIS 660
 DB 601 GNDVYWDDEKECFRTVASAVNFYALHPPIIPNSGNGIHLYKKNRDSMADEHAENDLIS 660
 QY 661 DENVDQELAEAEAAQAQREMTIOHVLFPSMRPLFLPKPKSMATDGFVQVASEKLYKI 720
 DB 661 DENVDQELAEAEAAQAQREMTIOHVLFPSMRPLFLPKPKSMATDGFVQVASEKLYKI 720
 QY 721 PERC 724
 DB 721 PERC 724

RESULT 2

ID AAE08710 standard; Protein: 737 AA.

AC AAE08710;

DT 15-NOV-2001 (first entry)

DE Arabidopsis thaliana PMS2 protein homologue MLH.

KM Hypermutable plant; dominant negative allele; mismatch repair gene;

KW MMR; cell line generation; PMS2; AtMLH.

OS Arabidopsis thaliana.

XX WC200161012-A1.

XX 23-AUG-2001.

XX 28-DEC-2000; 2000WO-US35397.

XX 18-FEB-2000; 2000US-0183333.

XX (NICO/) NICOLAIDES N C.

XX (GRAS/) GRASSO L.

XX (SASS/) SASS P M.

XX (KINZ/) KINZLER K.

XX (VOGE/) VOGELSTEIN B.

XX Nicolaides NC, Grasso L, SASS PM, Kinzler K, Vogelstein B;

XX WPI; 2001-529913/58.

XX Example 1; Page 57-59; 72pp; English.

CC The invention relates to a method for generating hypermutable cell.
 CC The method involves introducing into a plant cell a polynucleotide
 CC comprising a dominant negative allele of a mismatch repair (MMR) gene.
 CC The method is useful for generating hypermutable plants, new cell lines
 CC and plant varieties. This is particularly useful for agriculturally
 CC important crops. The method is also useful for generating crop plants
 CC with new output traits and plant cells exhibiting new biochemicals for
 CC commercial use. The present sequence is Arabidopsis thaliana (At)

CC MLH protein. This sequence is a homologue of MMR protein, PMS2.
 YY

SD Sequence 737 AA;

Query Match	Score	DB	Length
67.58	2505	22	737

Matches 482; Conservative 111; Mismatches 129; Indels 4; Gaps 3;

QY	2	DEPPRGCGGCGEPPIRRLREESVYVNIAGAEOYORSSAVKLEIENSIDAGASSVAV	61
Dd	13	EEESPATTIYPREPKIQRLEESVYVNIAGAEOYIQRVSAVKELVENSIDADSSISVAV	72
QY	62	KDGLKLIQVSDDHGIRFEDLALICERHTTSKLSAYEDLOTIKSMGRGALASMTYVG	121
Dd	73	KDGLKLIQVSDDHGIRREDLPILICERHTTSKLTREDFLSLSMGREGALASMTYVA	132
QY	122	HAVTITTEBOLHRYRYSDGWENEPKCAAVKGTQVAVENLFTINMAKKTLONSND	181
Dd	133	HAVTITITKGOIHGYRVSRYDGVHEHPKCAAVKGTQIVAVENLFYMIARRKTLQNSAD	192
QY	182	DYKPIVDIPISEFAVHHINMFPSCKHCAHMANADHASSTSSRLDAIRSVGASVVRDIEI	241
Dd	193	DYGIIVDLISMAIHHYNNVSPCKHCAVADHASVSPSRDLSIRSVGVSAKINLMK	252
QY	242	KVSEYEDADSIFFKMDGYISANANYAKKITMILFINDRJYCTALKRAIEFVSATLPOAS	301
Dd	253	EVSSEDCSGCGFDMEGEITSNMNYAKKITIYLFINDLVCSALKRAIELYVAATLPKAS	312
QY	302	KPIFYMSIHLPEHADVNIHPTKKEVSLNQBRIETIRNAIEBKLMNSTTRIIFOTQAL	361
Dd	313	KPIFYMSINLPREHVDINHPTKKEVSLNQBRIETIEMIOSEVEVKLTINADTRTFOEKV	372
QY	362	NLSGIAQANPOKOKVSEASMSGSGTKOKIPVSOVVRDPNPNPSGRLHTYHHGSSNLEK	421
Dd	373	EYIO-STUTSOKSDSPVYOKPBGSKITOKVYNNKVRKVRDSDSPGRLLAFIOPRQSPDK	431
QY	422	FDLVS-VARNVRSRRNOKDAGDSSRHELVEIDSSHFPLDLYVKNCTYVGADAEFAL	480
Dd	432	VSSLVSRRSVRQRNRPKRETDADSSVDELIAGVDSCHPMLETFVRCTYVGADADVEAL	491
QY	481	IQHTRILYLVAVVNIISKELMIOQLCRFGNPAIQLSEPPARLOELVMAKKBDEL--MSD	538
Dd	492	VOYTHIYLAIVVNLISKELMYOQTLRRFAHFAIQLSDPARPASELJILALKEBDLPDGN	551
QY	539	EKDDEKLEIAEVNTEIILKENAMENEFSHIIDODGKLTFLPYVLDDYPPDMRLRPEVL	598
Dd	552	TKDULKERIAMNTEILKERAMLEEFVSHIDSSANLSTRIPVILDDYTPDMRVPEFL	611
QY	599	ALGNDVYTWDEKECFRTIVASAVNGFYALHPILIPNPSGNGIHLHYKKRSDMADEHENDL	658
Dd	612	CLGNDVVEREDKSCFOGVSAALIGFYAMHPRLIPNPSGDDIOTYRKSGEBSQKSDLEGN	671
QY	659	ISDENVDQELLAEAAAMAQRENTIQHVLEPFSMKRFLKPPKSMATDGTFOVYASLEKLY	718
Dd	672	YDMEDNIDQDILLDAENAMAOREWSIQHVLEPFSMKRFLKPPASMASNGTIVYASLEKLY	731
QY	719	KIFPERC 724	
Dd	732	KIFPERC 737	

pollen formation; base pair mismatch.

OS *Arabidopsis thaliana*.

PN WO200224890-A2

PD 28-MAR-2002.

PF 18-SEP-2001; 2001WO-US29088.

PR 18-SEP-2000; 2000US-233124P.

PA (PION-). PIONEER HI-BRED INT INC

PI Mahajan PB;

DR WPI; 2002-416283/44

XX

PT efficiency of targeted gene mutation or homologous recombination in a

plant and for generating plants with
XX

The invention relates to isolated rice *MLH1* orthologue nucleic acids. The nucleic acid is useful for increasing the efficiency of targeted gene mutation or homologous recombination in a plant, by transforming a plant with expression cassette comprising the nucleic acid linked to a chemical inducible promoter, transforming the plant with nucleic acid comprising a sequence having a desired mutation or a sequence to be homologously recombined, where the transformation occurs in the presence of chemical compound capable of inducing the promoter and the plant's cellular mismatch repair system is inhibited and selecting the transformed plants that contained the mutation or homologously recombined nucleotide sequence. The plant cellular mismatch repair system is inhibited through the use of transposon tagging of an *MLH1* gene, sense- and antisense-suppression of an *MLH1* gene, antibody binding to an *MLH1* polypeptide or its variant, and targeted mutagenesis of specific amino acid residues encoded by an *MLH1* gene. The nucleic acid is also useful for producing reversible male sterility in a plant, by transforming a plant with an expression cassette comprising a *lexA* DNA binding site embedded in a tissue-specific promoter that drives expression in the plant operably linked to the nucleic acid when expressed disrupts pollen formation or function through inhibition of the plant's cellular mismatch repair system, transforming the plant with a second expression cassette comprising a nucleotide sequence encoding a *lexA* repressor protein operably linked to a chemically-inducible promoter that drives expression in the plant, and exposing the plant to a compound capable of inducing the chemical-inducible promoter, to induce expression of *lexA* repressor protein. The tissue-specific promoter is an anther-specific promoter and the chemical-inducible promoter is a herbicidal safener. The polypeptide encoded by the nucleic acid is useful for detecting, locating, or removing a base pair mismatch (SNP). The present sequence is Arabidopsis *Thaliana MLH1* protein.

Sequence **737 AA;**

query match	67.58;	Score 2505;	DB 23;	Length 737;
-------------	--------	-------------	--------	-------------

Matches 482; Conservative 111; Mismatches 129; Indels 4; Gaps 3;

Qy	2	DEPPRGCGCGEPPRIPIRLIEESVYNNR	LAAGEVIORPSNAKTELIEENSLAGASVSVAV	61
Db	13	EESPAATIVPREPKIQRLIEESVYNNR	LAAGEVIORPSNAKTELIEENSLAGASVSVAV	72
Qy	62	KDGGILKILQVSDDHGIRFEDALITLCRRHTSK	SLAEDQITKSGFPEALASMTYNG	122
Db	73	KDGGILKILQVSDDHGIRFEDALITLCRRHTSK	SLAEDQITKSGFPEALASMTYNG	133
Qy	122	HATVTTITTEGOLHGIRVSYRDKGVNENEPK	PCAAVKGTVAMVENVNNAVARKTLQNSD	181
Db	133	HATVTTITTEGOLHGIRVSYRDKGVNENEPK	PCAAVKGTVAMVENVNNAVARKTLQNSD	192

[illegible]

Discriminating proliferating from non-proliferating cells in tissue using antibodies specifically immuno-reactive with mismatch repair protein, esp. human MSH2

Disclosure: Page 23-25; 37pp; English.

The sequences given in AAW09034-36 represent the human mismatch repair proteins, hMSH1, hMLH1 and hPMS2. In the method of the invention, these proteins were identified by reaction with an antibody (Ab) specific for them, therefore discriminating proliferating from non-proliferating cells. The method may be used for monitoring the effectiveness of anti-cancer therapy in neoplastic tissue, by comparing the amount of Ab-Ag complexes in the sample with an amount determined at an earlier time, in which a reduction in the amount indicates an effective therapy. The Ab are especially specifically immunoreactive with the MSH2 mismatch repair gene, which is 1 of at least 4 genes encoding proteins involved in the repair of mismatched nucleotides following DNA replication or repair. Mutations in the MSH2 gene contribute to the development of sporadic colorectal carcinoma, while germline MSH2 mutations are responsible for approx. 50% of inherited, non-polyposis colorectal carcinoma (HNPCC). Since MSH2 is ubiquitously expressed, development of other cancers are also susceptible to alterations in MSH2.

Query Match	39.3%;	Score 1457;	DB 18;	Length 752;
Best Local Similarity	39.7%;	Pred. No. 2.5e-111;		
Matches 311;	Conservative 146;	Mismatches 212;	Indels 114;	Gaps 11;

[illegible]

QY 602 NDYTWDEKECFRTVASANGNYALHPILPNPNSGNGILYKKNRDSMADEHAENDLISD 661
DB 662 TEVNMDEKECFESLSKECAMFYSI-----RKQYISE 693
QY 662 ENDVODELAEAAAREMTIOHVLFPSPMLFLKPKSMATDGTGFVOVASLEKLYKIF 721
DB 694 ESTLSGQO-SEVPGSW---KMTVEHIVYKALRSHILPKHFTEDGNILQIANLPDLKYVF 749
QY 722 ERC 724
DB 750 ERC 752

RESULT 5

AAR75785
ID AAR75785 standard; Protein; 756 AA.

AC AAR75785;

DT 04-MAR-1996 (first entry)

DE Human wild type MLH1, a Muhl homologue.

KM bMLH1: wild type: Muhl homologue; cancer diagnosis; mismatch repair;
tumour; susceptibility; mutation detection.

OS Homo sapiens.

PN WO9516793-A1.

PD 22-JUN-1995.

PF 16-DEC-1994; 94WO-US14746.

PR 09-DEC-1994; 94US-0352902.

PR 17-DEC-1993; 93US-0168877.

PR 08-MAR-1994; 94US-0209521.

PA (DAND) DANA FARBER CANCER INST INC.

PA (UYOR-) UNIV OREGON HEALTH SCI.

PI Baker SM, Bollag RJ, Bronner CE, Kolodner RD, Liskay RM;

DR WPI, 1995-231583/30.

DR N-PSDB; AAO90814.

PS Claim 33; Fig 3; 168bp; English.

CC AAO90814 encodes AAR75785 the wild type bMLH1, a Muhl homologue. A
CC mutation in a bMLH1 or hPMS1 nucleic acid isolated from a subject,
CC can be detected by comparing it with an analogous segment of the
CC above wild type allele. This method can be used to diagnose cancer
CC susceptibility, or to identify and classify a DNA mismatch-repair
CC defective tumour.

XX Sequence 756 AA:

Query Match 39.1%; Score 1452; DB 16; Length 756;

Best Local Similarity 39.6%; Pred. No. 6.5e-111;

Matches 311; Conservative 145; Mismatches 214; Indels 116; Gaps 11;

QY 18 IRLLESYVNRITAGEVIOFPSAVKELIENSADAASSVAVKGGKLTQVSDGHS 77
DB 8 IRLDSTVNRITAGEVIOFPSAVKELIENSADAASSVAVKGGKLTQVSDGHS 67
QY 78 IREDIALICERHTTSKLSAYEDLOTIKSMGFGEALASTVYGVHTVTTTREGOLHGV 137
DB 68 IREDLDIYCERHTTSKLSAYEDLOTIKSMGFGEALASTVYGVHTVTTTREGOLHGV 127

QY 138 VSTRDGVMEKEKPCQAVKGTQVMEVNFNMVARKKTLQNSDDYPKIYDTSRAVHH 197
DB 128 ASYSDCKLAPKAPKPCAGNOGTQITVEDLFYNATRRKALKNPSEEGKILEVVGGRSVNH 187
QY 198 INTFESCRHGANRAVHASTSSRLDAIRSVYGASVVDLIEIKSYEDAASITKMG 257
DB 188 AGISFVKKOGETVADVRLTPNASTVDNIRSTIGNAVSRLEITIGGEDKTLA---TKMG 244
QY 258 YISNANYVAKKITMILFINDRLVDCATLAKRAIEFVVSATLPQAKRPFIYMSIHLPSEHD 317
DB 245 YISNANYSVKCTFLFIHNRIVESTSLKRAIETVYAALPKRTHPFIYLSLEISQND 304
QY 318 VNIHPYKKEVSLNQGRIETIRNAIEEKLMSNTTRIFOTQALNLSGIAQANPQ---KDK 375
DB 305 VVNHPTKHEVHFLHESILSERVOQHIESKLSGSSNRMVFTQTL-LPGLAGPGEVWKST 363
QY 376 VSEASMSGSTKOKIPVSOVRRDPNRPNSRLTFYHGGSSNL-----KKKDIYS 426
DB 364 TSLTSSSTSGSDKVAHOGVRTDSRQ--KIDAFIQPLSKPLSOPQALVTEDKTDISS 421
QY 427 VRNVVR-----SRNQKDAGLS-----444
DB 422 GRARQODEMELPAPAEVAANKQSLGDTTKTSEMSEKRGPTSSNPKRHRDSDVEM 481
QY 445 -----SRHELY-----EIDSFHGLDIYKCTYGLADEAPALI 481
DB 482 VEDDSRKEWTAACTPRRRIINLTSLVSLQELNEOGEVLEKMLAHNSFVGCYVPMWALA 541
QY 482 OHNTRLIYVNVNISKELWQALCRPGNENAIOLSEPAIOLILYMAKDDLMODEX 541
DB 542 OHOTKLYLNTTKLSELEFYQIILYDPAANGVLRLESPALFLAMIALDSPSGWTED 601
QY 542 DEKLEIAEVNTEILKENAEMINEYFSIHIDODGKLTRLPAVLDOYTPDMRLPEFYALG 601
DB 602 GREGLAEVLYVEELKKKAEMADYFSEIDEENLGLPLIDNYVPLEGLFIFILRLA 661
QY 602 NDYTWDEKECFRTVASANGNYALHPILPNPNSGNGILYKKNRDSMADEHAENDLISD 661
DB 662 TEVNMDEKECFESLSKECAMFYSI-----RKQYISE 693
QY 662 ENDV---DELLAEAAAREMTIOHVLFPSPMLFLKPKSMATDGTGFVOVASLEKLY 718
DB 694 ESTLSGQOSEVPGISPSW---KMTVEHIVYKALRSHILPKHFTEDGNILQIANLPDL 750
QY 719 KIFERC 724
DB 751 KYPERC 756
RESULT 6
AAR76071
ID AAR76071 standard; Protein; 756 AA.
AC AAR76071;
DT 15-JAN-1996 (first entry)
DE Human mismatch repair pathway protein, Mhl.
KM Mismatch repair; MSH2; primer; identification; defect; alteration;
KW cancer; tumour; vaccine.
OS Homo sapiens.
PN WO9514085-A2.
QY 26-MAY-1995.
DB 17-NOV-1994; 94WO-US13385.
QY 13-JUN-1994; 94US-0259310.
PR 17-NOV-1993; 93US-0154792.
PR 07-DEC-1993; 93US-0163449.

CC Penicillin G, Erythromycin and Clavulanic acid, by biotransformation.
 CC Dominant negative alleles of the MMR gene are useful for producing higher
 CC quantities of recombinant polypeptides. The present sequence represents
 CC a human MLH1 protein.

XX Sequence 756 AA;

Query Match 39.1%; Score 1452; DB 22; Length 756;
 Best Local Similarity 39.6%; Pred. No. 6,5e-111;
 Matches 311; Conservative 145; Mismatches 214; Indels 116; Gaps 11;

QY 18 IRLLEESVNRIRIAGVYIQRPSAVKELIENSIDAGASSVAVKOGGLKLIQVSDGHC 77
 DB 8 IRLDELTVNRIAGVYIQRPSAVKELIENSIDAGASSVAVKOGGLKLIQVSDGHC 67
 QY 78 IRLDELTAICERTHTSKLSAYEDLQITKSGFGEALASMTYGVHTVTITTEGQLHGVR 137
 DB 68 IRLDELTAICERTHTSKLSAYEDLQITKSGFGEALASMTYGVHTVTITTEGQLHGVR 127
 QY 138 VSTRDGMENEPKPCAAVKGTOYVNEFLFYNNVARKKTLQNSDDYKTIYDFISRFVHH 197
 DB 128 ASYSDGKLKAPKPCAGNOGTQITVEDLFYNNVARKKTLQNSDDYKTIYDFISRFVHH 187
 QY 198 INTFECRKHGARNADYHSAHSTSSRLDAIRSVYGVASVNDLITIKSYEDAAISIRKMG 257
 DB 188 AGISFVKKOGETVADYVRLPNASTVDNIRSIIPGNAVRELIIGCEDKTLA---FKMG 244
 QY 258 YISNAYVAKKITMLEINDRLVDCTALKRAIEFVSATIPQASKPIYAKSHLPSEHYD 317
 DB 245 YISNAYVAKKITMLEINDRLVDCTALKRAIEFVSATIPQASKPIYAKSHLPSEHYD 304
 QY 318 VNIHPKKEVSLNDRRIETINAEIKLMSNTTRIFQTOALNLSGLAAMPQ--KDK 375
 DB 305 VNIHPKKEVSLNDRRIETINAEIKLMSNTTRIFQTOALNLSGLAAMPQ--KDK 363
 QY 376 VSASAGSGTSSKQITVYSDVNRTPNPSGRITTYHGGSSNL-----EKKEDLYS 426
 DB 364 TSTSSSTSGSSDKYVAHQMVRTDSRRO--KIDAFIQPLSKPLSOPQALVTFEDKTDISS 421
 QY 427 VRNVV-----SRNOKDAGDS----- 444
 DB 422 GRARODEEMLELPAPEVAANKNOSLEGDTTKTSEMSEKRGPTSSNPKRRHEDSDVEM 481
 QY 445 ---SHELLV-----EIDSFHPGLDIYVKNCTYVGLADEAFALI 481
 DB 482 VEDDSREKMTAACTPRRIINTLSVLSLQEIENEGHEVLRKMLHNSFEVGVNPOWALA 541
 QY 482 QHNTRIYVNVVNIKELWQALCRFGNNAIQISEPARLQELWALKDDDELMSDEKD 541
 DB 542 QHNTRIYVNVVNIKELWQALCRFGNNAIQISEPARLQELWALKDDDELMSDEKD 541
 QY 542 QHNTRIYVNVVNIKELWQALCRFGNNAIQISEPARLQELWALKDDDELMSDEKD 601
 DB 542 DEKLEIAVNTTEIKENAEINNEYFISIHIDODKLTIRLPVLDQYTPMDRLDFEYVATG 601
 QY 602 GREGLAETIVEFLKKAKEADYFSLIEDENLIGPLIDINNYPPLEGGLFIILIRLA 661
 DB 602 GREGLAETIVEFLKKAKEADYFSLIEDENLIGPLIDINNYPPLEGGLFIILIRLA 661
 QY 602 NDVTWDEKCEFTVAVASVGNFYALHPILPNPNSGNGIHLKKNRDSMADEHAENDLISD 661
 DB 662 TEVNMDEKCEFTVAVASVGNFYALHPILPNPNSGNGIHLKKNRDSMADEHAENDLISD 661
 QY 662 TEVNMDEKCEFTVAVASVGNFYALHPILPNPNSGNGIHLKKNRDSMADEHAENDLISD 693
 DB 662 TEVNMDEKCEFTVAVASVGNFYALHPILPNPNSGNGIHLKKNRDSMADEHAENDLISD 693
 QY 662 ENDV---DDELLAEARAAQORWTIQHVLFPSPKRLFLPKPKMAWDGTFVQVASELKY 718
 DB 694 ESTLSGQOOSVPGSIPNSW---KMTVEHLYVKALRSHIILPKHFTEDGNILQIANLPDLX 750
 QY 719 KIPERC 724
 DB 751 KYFERC 756

RESULT 8

AA63956
 ID AA63956 standard; protein; 756 AA.
 XX
 AC AA63956;

XX 29-OCT-2001 (first entry)
 XX Amino acid sequence of human mismatch repair protein MLH1.
 DE MLH1, mismatch repair gene; MMR gene; hypermutable yeast.
 KW Homo sapiens.
 OS Homo sapiens.
 XX MO200162945-A1.
 XX 30-AUG-2001.
 PD 21-FEB-2001; 2001MO-US05447.
 XX 23-FEB-2000; 2000US-0184336.
 XX (UYJO) UNIV JOHNS HOPKINS.
 PA (NICO) NICOLAIDES N C.
 PA (SASS) SASS P M.
 PA (GRASO) GRASSO L.
 PA (VOGE) VOGELSTEIN B.
 PA (KINZ) KINZLER K W.
 XX Nicolaides NC, Sass PM, Grasso L, Vogelstein B, Kinzler KW;
 PI N-PSDB: AAH75044.
 DR N-PSDB: AAH75044.
 XX

PT Making hypermutable yeast that exhibit novel selected output traits for
 PT commercial applications, comprises introducing polynucleotide
 PT containing dominant negative allele of mismatch repair gene
 PS Disclosure; Page 43-44; 60pp; English...

CC The present sequence represents human MLH1. MLH1 is a mismatch repair
 CC (MMR) gene. The specification describes a method for making a
 CC hypermutable yeast, comprising introducing a polynucleotide containing
 CC a dominant negative allele of a mismatch repair (MMR) gene, into a
 CC yeast, whereby the cell becomes hypermutable. The method is useful
 CC to create desirable output traits for commercial applications, using
 CC dominant negative alleles of mismatch repair proteins. The hypermutable
 CC yeast is useful for production, biocatalysis, bioremediation and drug
 CC discovery. It is also useful in genetic screens for the direct selection
 CC of variant subclones that exhibit new output traits. The hypermutable
 CC yeast is also useful in the manufacturing industry for the generation
 CC of new biochemicals, for detoxifying noxious chemicals from by-products
 CC of manufacturing processes or those used as catalysts, for remediation
 CC of toxins present in the environment including polychlorobenzenes, heavy
 CC metals and other environmental hazards for which there is a need to
 CC remove them from the environment. The yeast is also useful for producing
 CC increased quantity or quality of protein or non-protein therapeutic
 CC molecule e.g., Penicillin G, Erythromycin and Clavulanic acid, by
 CC biotransformation.

XX Sequence 756 AA;

Query Match 39.1%; Score 1452; DB 22; Length 756;
 Best Local Similarity 39.6%; Pred. No. 6,5e-111;
 Matches 311; Conservative 145; Mismatches 214; Indels 116; Gaps 11;

QY 18 IRLLEESVNRIRIAGVYIQRPSAVKELIENSIDAGASSVAVKOGGLKLIQVSDGHC 77
 DB 8 IRLDELTVNRIAGVYIQRPSAVKELIENSIDAGASSVAVKOGGLKLIQVSDGHC 67
 QY 78 IRLDELTAICERTHTSKLSAYEDLQITKSGFGEALASMTYGVHTVTITTEGQLHGVR 137
 DB 68 IRLDELTAICERTHTSKLSAYEDLQITKSGFGEALASMTYGVHTVTITTEGQLHGVR 127
 QY 138 VSTRDGMENEPKPCAAVKGTOYVNEFLFYNNVARKKTLQNSDDYKTIYDFISRFVHH 197
 DB 128 ASYSDGKLKAPKPCAGNOGTQITVEDLFYNNVARKKTLQNSDDYKTIYDFISRFVHH 187

DB	Query	Score	39.1%	Score 1452	DB 23	Length 756
DB	Best Local Similarity	39.6%	Pred. 6.5e-111			
XX	Matches 311; Conservative 145; Mismatches 214; Indels 116; Gaps					
XX	Sequence 756 AA:					
DB	18 IRLRELSVYNNRIAGAEVYQVRSSAVKELIENSLDAGASSVVAVDGKLITQVSDGHC 7					
DB	8 IRLRDETVYNNRIAGAEVYQVRSSAVKELIENSLDAGASSVVAVDGKLITQVSDGHC 7					
DB	78 IREEDALICERHTYTSKLSAEEDLOTIKSGFGEALSMYVGVYVYTTTBEQLHGR 13					
DB	68 IREEDALICERHTYTSKLSAEEDLOTIKSGFGEALSMYVGVYVYTTTBEQLHGR 13					
DB	138 VSTRGVQVNEPKCAAVKGTQVAVENLEFVNNVARKTLQNSNDDEPKYVDFISREFAVH 19					
DB	128 ASYSGKTKAPKPCACAOOGTQTVDELFYNIATRKALKNPSEYKGLLEVGRASVYN 18					
DB	198 IAVTSCKRHANADVADHSATSSRDIAIRSYGASVYRDLIKVSEEDPADSIFKMG 25					
DB	188 AGISRVAKQGEIYADVTRLPNASTYVDNTRISFGNVSRELIEICEDKTLA--FKMG 24					
DB	256 YISNANYAKTITLFIINDRLVDTALRALEFYYSATLPOASKPTIYSHLSEHND 31					
DB	245 YISNANYAKTITLFIINDRLVDTALRALEFYYSATLPOASKPTIYSHLSEHND 31					
DB	318 VNIHTKKEYSLNOEIIETITRNAIEELKMSNTTRIFETQALNLSGIAQANPQ--KDK 37					
DB	305 VNIHTKKEYSLNOEIIETITRNAIEELKMSNTTRIFETQALNLSGIAQANPQ--KDK 37					
DB	376 VSEASMGSGTRKIPQSVQVTRDPNRSGLRHLYWQGOSSNL-----EKKDLVS 42					
DB	364 TSLTSSSTSGSSKYVAHQMYVTRDSNEG-KLDAFLQPLKPLSSOPALVTEDTIDSS 42					
DB	427 VRNVYR-----SRNQGDAADLS----- 44					
DB	422 GRANQOQDEMLELPADAEVAANOSLEDGTTKSGSESKRGPPTSNPKRRHREDSDVM 48					
DB	445 -----SRDELIV-----EIDSPFPGLLDLYKCTYVGLADARFALI 48					
DB	482 VEDDSREKMTACTPERRINTSLVLSQELINQGEVLYREMLHNSHVSFGVQNMAL 54					
DB	482 OHNTRILYVAVNISKELMAYQALCREGNEFNAIQLSEPAQLDELVALKDELMSDKD 54					
DB	542 QHOKRIYLLNTKLSBELFYQILTYDFANFGEVTLSEPAFLDPMALALDPSGWTED 60					
DB	542 DEKLEIAEVNTEILENEMENEFYSHIIDQGLTRPLRVYVDQTYTTPMDRLPEFVALG 60					
DB	602 GREGIAETIYVEFLKKKMLADIFSLDEBEGNLGLPLLDNNVYVPLEGPIFLITLA 66					
DB	602 NDVTWDEKERTVASAVNGEVYALHPPIIPNSGNGIHLYKKNRNSMADFAFNNDISD 66					

DB 662 TEVNMDEKECFESLSEKAMFYSI-----RKQYISE 693
QY 662 ENDV---DDELLAEAAEAQAQREMTIOHVLFPMSRLFLAPPKSMATDGTFFVQVASELEKLY 718
DB 694 ESTLSGQSEVPSPINPSW---KWTVEHIVYKALRSHIILPPKHFTEBDGNILQIANLPDLY 750
QY 719 KIFERC 724
DB 751 KIFERC 756

RESULT 10
AAU98778
ID AAU98778 standard; Protein; 756 AA.
AC AAU98778;
XX
XX
XX 24-SEP-2002 (first entry)
DE Human Mult. homologue, MLH1.
XX
XX Human; DNA mismatch repair enzyme; MLH1;
KW Mut. L. homologue; dominant negative mutation; antibacterial;
KW hypermutable cell; post meiotic segregation increased; PMS1; PMS2;
KW Mult. homologue 2; MSH2; MLH2; antimicrobial.
XX
OS Homo sapiens.
XX WO200238750-A1.
XX
XX 16-MAY-2002.
XX
XX 07-NOV-2000; 2000WO-US30587.
XX
XX 07-NOV-2000; 2000WO-US30587.
XX
XX 07-NOV-2000; 2000WO-US30587.
XX
XX (MORP-) MORPHOTEK INC.
XX
XX Grasso L, Nicolaides NC, Sass PM;
PI WPI: 2002-508210/54.
XX
XX N-PSDB; ABR66091.
DR
XX
XX Making mammalian cell hypermutable for obtaining a mammalian cell that
PT is resistant to selected microbe by introducing polynucleotide
PT comprising dominant-negative allele of mismatch repair gene into
PT mammalian cell
XX
XX
XX Disclosure; Page 18; 68pp; English.

The invention relates to making (M1) a mammalian cell hypermutable comprising introducing a polynucleotide with a dominant-negative allele of a DNA mismatch repair gene (e.g. post meiotic segregation increased 1 and 2 (PMS1 and PMS2), Mult. homologue 2 (MSH2) and Mult. homologue 1 and 2 (MLH1, 2) into the mammalian cell which is hypermutable. Also included is a homogenous composition comprising a cultured, hypermutable, mammalian cell with a dominant negative allele of mismatch repair gene. (M1) is useful for making a mammalian-cell hypermutable. The composition is useful for obtaining a mammalian cell that is resistant to a selected microbe. (M1) involves growing a culture of mammalian cells with a dominant negative allele of a mismatch repair gene, exposing the cells to the selected microbe, and selecting the mammalian cell resistant to the selected microbe. The hypermutable cell is resistant to gram-negative and gram-positive microbe, protozoan, bacterium or fungi. The microbial resistance is selected by isolating and testing conditioned medium from the hypermutable cell. The composition is also useful for obtaining a cell comprising a mutation in a gene encoding an antimicrobial activity. (M1) involves growing a culture of mammalian cells with the gene encoding the antimicrobial activity, and a dominant negative allele of a mismatch repair gene, selecting a cell comprising the antimicrobial activity, and determining whether the gene comprises a mutation. The cell is examined to determine whether the gene comprises a mutation by analysing a

CC nucleotide sequence of the gene or mRNA transcribed from the gene, a
CC protein encoded by gene or its phenotype. The composition
CC is useful for identifying antimicrobial agents, microbe-specific toxic
CC molecules, and for producing new phenotypes of the cell. (M1) is useful
CC for creating genetically altered antimicrobial molecules, and also for
CC creating cell lines that manufacture antimicrobial molecules for use in
CC large scale production of antimicrobial agents for clinical studies.
CC (M1) is also useful in cell lines that express known antimicrobial
CC agents to enhance the biochemical activity of the antimicrobial agent.
CC The present sequence represents human MLH1.
XX
XX
XX Sequence 756 AA:
Query Match 39.1%; Score 1452; DB 23; Length 756;
Best Local Similarity 39.6%; Pred. No. 6.5e-111;
Matches 311; Conservative 145; Mismatches 214; Indels 116; Gaps 11;
QY 18 IRLLESYVNRIRAAAGEYIOPSSAVKELINSIDAGASSVAVKDGGLKLIQVSDGHHG 77
DB 8 IRLDETIVNRIRAAAGEYIOPSSAVKELINSIDAGASSVAVKDGGLKLIQVSDGHHG 67
QY 78 IRLDETIVNRIRAAAGEYIOPSSAVKELINSIDAGASSVAVKDGGLKLIQVSDGHHG 137
DB 68 IRLDETIVNRIRAAAGEYIOPSSAVKELINSIDAGASSVAVKDGGLKLIQVSDGHHG 127
QY 138 VSYRDGEMENEPKCAVKTQVAVENLFYVMAVARKKTIQNSDDPKIVDFISRPVHH 197
DB 128 ASYSDGKLKAPKPCAGNQTQITVEDLFYVMAVARKKTIQNSDDPKIVDFISRPVHH 167
QY 198 INTFSCRRKGNARADYHASTSRDLAISVYGASVYRLIEKYSIEDAADSIFMDG 257
DB 188 AGISFVYKGGETVADRTLPNASTVDNIRISIFGNASRLEIGCECDKTLA--FEMNG 244
QY 258 YISNANYVAKKITMILFINDRLVDCFKRAIFEVYSATLPQASKPPIYVSIHLPSEHV 317
DB 245 YISNANYVAKKITMILFINDRLVDCFKRAIFEVYSATLPQASKPPIYVSIHLPSEHV 304
QY 318 VNIHPTKREYSLNQEIRIETIRNAIEEKLNSNTTRIFQTLALNSGIAQNPQ--KDK 375
DB 305 VNIHPTKREYSLNQEIRIETIRNAIEEKLNSNTTRIFQTLALNSGIAQNPQ--KDK 363
QY 376 VSEASMGSGTKSOKIPYSONVTRDPRNSGRLHTYHGGSSNL-----EKKEDLV 426
DB 364 TSLTSSSTSGSDKYVAHQVARTDSREQ--KLAFLOPLSKPLSPQATVTEDEKTDISS 421
QY 427 VRVNVR-----SRNOKDAGDLS----- 444
DB 422 GRARQDEEMLELPAPAEVAKNOSLEGDTKGTSESEKRGPTSSNRRHREDSDEM 481
QY 445 -----SRHELLY-----EIDSFHFGGLDVKVKNCTVYGLADEAFALI 481
DB 482 VEDDSKREMAACTPRRIINISVLSLOEINOGHEVIREMHNHNSPFCVNPQALA 541
QY 482 QHNTRILYVNVNISKELMTOALCRGNPNAIOLSEPADLOELLYVALKDELMSKED 541
DB 542 QHOFKTYLNTTTLSELEFQOILYDFANGVIRLSEPADLOELLYVALKDELMSKED 601
QY 542 DEKLEIAEVNTELLKEAENINEFSTHIDODKTRPLVLIQVYPMDBMRLEFVALG 601
DB 602 GPEGLAEVTEVLEKKAEMADYFSEIDEGELGLPLIDVYPPLEGLPFIIRLA 661
QY 602 NDVTWDEKECFEFTVASAVNGFYALHPILPNPSGNGIHLKKNRSMADHAEENDLISD 661
DB 662 TEVNMDEKECFESLSEKAMFYSI-----RKQYISE 693
QY 662 ENDV---DDELLAEAAEAQAQREMTIOHVLFPMSRLFLAPPKSMATDGTFFVQVASELEKLY 718
DB 694 ESTLSGQSEVPSPINPSW---KWTVEHIVYKALRSHIILPPKHFTEBDGNILQIANLPDLY 750
QY 719 KIFERC 724
DB 751 KIFERC 756

RESULT 11
AAG75536 standard; Protein: 775 AA.
ID AAG75536 standard; Protein: 775 AA.
AC AAG75536;
DT 03-SEP-2001 (first entry)
DE Human colon cancer antigen protein seq ID NO:6300.
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 3.
OS Homo sapiens.
PN WO200122920-A2.
PD 05-APR-2001.
PF 28-SEP-2000; 2000WO-US26524.
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX (HUMA-) HUMAN GENOME SCI INC.
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
PI WPI: 2001-235357/24.
DR N-PSDB; AAH34941.
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
PS Claim 11; Page 7756-7760; 9803pp; English.
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens (P), where
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patient's own production of P.
CC Additionally, N may be used to produce the colon cancer-associated P,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX Sequence 775 AA:
Query Match 39.1%; Score 1452; DB 22; Length 775;
Best Local Similarity 39.6%; Pred. No. 6,8e-111;
Matches 311; Conservative 145; Mismatches 214; Indels 116; Gaps 11;
OY 18 IRRLEESVYNIAGETIYORSSAVKELIENSIDAGASSVAVKDGKLIQVSDGHR 77
DB IRRLEESVYNIAGETIYORSSAVKELIENSIDAGASSVAVKDGKLIQVSDGHR 77
OY 78 IRRLEESVYNIAGETIYORSSAVKELIENSIDAGASSVAVKDGKLIQVSDGHR 86
DB IRRLEESVYNIAGETIYORSSAVKELIENSIDAGASSVAVKDGKLIQVSDGHR 86
OY 87 IRRLEESVYNIAGETIYORSSAVKELIENSIDAGASSVAVKDGKLIQVSDGHR 137
DB IRRLEESVYNIAGETIYORSSAVKELIENSIDAGASSVAVKDGKLIQVSDGHR 137
OY 138 VSYRDCVNEKPKCAAVGTQVMEVLEIYVARKKTLQNSDDYDKIYDIFSPFAVH 197
DB VSYRDCVNEKPKCAAVGTQVMEVLEIYVARKKTLQNSDDYDKIYDIFSPFAVH 197

RESULT 12
AAR79008 standard; Protein: 756 AA.
ID AAR79008 standard; Protein: 756 AA.
AC AAR79008;
DT 26-MAR-1996 (first entry)
DE Human DNA repair protein hMLH1.
KW DNA repair protein; hMLH1, hMLH2, hMLH3; therapy; cancer; vectors;
KW DNA synthesis; diagnosis; disease; multi4.
OS Homo sapiens.
PN WO9520678-A1.
PD 03-AUG-1995.
PF 25-JAN-1995; 95MO-US01035.
PR 23-AUG-1994; 94US-0294312.
PR 27-JAN-1994; 94US-0187757.
PR 16-MAR-1994; 94US-0210143.
XX (HUMA-) HUMAN GENOME SCI INC.
DB 147 ASYSDGLKAPKPCAGNOCTQITVEDLFFYNATRRKALKNPSEEGKILLEVGRYSVN 206
OY 198 INTFSCRRKGNRADVSHASTSSRLDAIRSYGASVVDLEITYSYDADASTFKKDG 257
DB 207 AGISFVYKOGELVADVRLPNASTVDNRSIFGNVNSVLEIECEDKTLA---FKNG 263
OY 258 YISNANYVAKKTIIMLEIFINDRLVDCALRAIEFYASATLPQASKPFIYMSIHSEHVD 317
DB 264 YISNANYVAKKTIIMLEIFINDRLVDCALRAIEFYASATLPQASKPFIYMSIHSEHVD 317
OY 318 VNIHPRKKEVSLINDRIETIRMAIEERKLNSTTRIFOTQALNLSGIAQANQ--KDK 375
DB 324 VNIHPRKKEVSLINDRIETIRMAIEERKLNSTTRIFOTQALNLSGIAQANQ--KDK 375
OY 376 VSEASMSGSGTKOKIPVSGMVTDPDPRNSGRLHYHWSQSNL-----EKKFDLVS 426
DB 383 TSLTSSSTSGSSDKVYAHQWRTDSRQ--KIDAFLOPLSKPLSSQPAIYTEDKTISS 440
OY 427 VNNVVR-----SRNOKDAGDLS-----444
DB 441 GRARQODEMELPAPAVAAKNOSLEDDTKGTSSEKRGPTSSNPRKRHEDSDVEM 500
OY 445 -----SRHELV-----EIDSSFRPGLDIYKNTYVGLADEAPALI 481
DB 501 VEDSRKEMTACIPRRRIINTSVLSLOEIRINQGEVLRMLHNSFVGCYNPQWALA 560
OY 482 QHNRLVLYVNVNISKELMVOALCREGNFNAIOLSEPAFLOELVWALKDELDMSDEKD 541
DB 561 QHOKLYLNTTKLSEELFYOLITDPANFVLRLESPALFDLMLADSPESGTEWD 620
OY 542 DEKLEIAVNTLEIKENAMINEFYSIHIDQDKLRLPVLDQYPPMDRLPEFLAG 601
DB 621 GPKGGLAEVLEFELKKAEMADYFSLDEEGNLGLFLIDNVPYPLEGPIFILRA 680
OY 602 NDVTWDEKCEFRVYASVGNFYALHPILPNPSNGGHIILYKKNDSMADEHENDLSD 661
DB 681 TEVWDEKCEFRVYASVGNFYALHPILPNPSNGGHIILYKKNDSMADEHENDLSD 661
OY 662 ENDV---DQELLAEANAKRBTIOHVLFPMSRLFKPPSMATDGTFOVASILEKTY 718
DB 713 ESTLSGQOSVPSIPSNW---KWVEHIVYALKSHILPPIPHFTEDGNIQLANLPOLY 769
OY 719 KIFERC 724
DB 770 KYFERC 775

XX	Adams MD, Fleischmann RD, Fraser CM, Fuldner RA;
PI	Haseltine WA, Kirkness EF, Rosen CA, Ruben SM, Wei Y;
XX	WPI; 1995-275461/36.
DR	N-PSDB; AA097525.
DR	
XX	
PT	Polynucleotide(s) encoding human mutL homologues, hMLH1, hMLH2 and
PT	hMLH3 - used for therapeutic treatment of, e.g. hereditary cancer
XX	
PS	Claim 8; Figure 1; 124pp; English.
XX	
CC	The polynucleotides described in AA097525-27 encode the human
CC	analogue of the prokaryotic mutL DNA repair gene. The polynucleotides
CC	they encode (AA097508-R79010) are used for therapeutic purposes e.g.
CC	in the treatment of cancer, esp. hereditary cancer. They may also
CC	be used for in vitro manipulation of DNA, synthesis of DNA and the
CC	manufacture of DNA vectors and in methods of diagnosing a disease or
CC	a susceptibility to a disease related to a mutation in the hMLH1, -2
CC	or -3 DNA repair genes.
XX	
XX	
Sequence	756 AA;
50	

Query Match	38.8%;	Score 1439;	DB 16;	Length 756;
Best Local Similarity	39.4%;	Pred. No. 7.8e-110;		
Matches 310;	Conservative 144;	Mismatches 216;	Indels 116;	Gaps 11

[illegible]

```
Db      662 TEVMMDEKECEFESELSKCAAFYSI-----RROYISE 693
QY      662 ENDY---DOELLAEAAAMAGREVTIOHLEPPSRMLFLPKPRSMATDGFVOVASLEKL 718
       1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      694 ESTLGGQOSEVPGSIPMS--KATVEHYIKALRSNHLPEPKHTEDGNILQLANLPDLX 750
QY      719 KIFERC 724
       1 : : : : :
Db      751 KVFERC 756
```

RESULT 13
ADDRESS 767

ID	standard; Protein; 664 AA.
ABB65767	

AC ABB65767;
yy

DT 26-MAR-2002 (first entry)
XY

Drosophila melanogaster polypeptide SEQ ID NO 24093.

KW Drosophila; developmental biology; cell signalling; insecticide; pharmacological

05 *Drosophila melanogaster*
yy

PN WO200171042-A2
YY

PD 27-SEP-2001
XY

PF 23-MAR-2001; 2001WO-US09231.
YY

PR 23-MAR-2000; 2000US-191637P.
 DP 11-MAY-2000; 2000US-0614150

XX
XX
PA (DEKE) DE CORD NY

XX	Venter, J.C.	Adams M
PI		

XX .
DB WBT: 2001-656960/75

DR N-PSDB; ABL09870.
XY

	New Isolated nuclei
PT	
DT	nucleus from Proctor

interactions :-

PS Disclosure; SEQ ID NO 24093; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from *Drosophila*. The invention is useful in developmental biology and in elucidating cell signaling and cell-cell interactions in higher and eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB101840-AB116175), expressed DNA sequences (AB101840-AB116175) and the encoded proteins (AAB571737-ABB720742).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 664 AA;

Query Match	34.0%	Score 1261	DB 22	Length 664
Best Local Similarity	36.9%	Pred. No. 3.3e-95		
Matches 270	Conservative 145	Mismatches 223	Indels 94	Gaps 11

QY 14 EPPRIRLEESVNR IAGEVIQRPSSAVKELIENSLDAGASSVSAVKDGLKLIQVSD 73

[illegible]

134 HGVRVSRDGVNENEPKCAAVKGTQVWENLFYNNVARKTLQNSNDYPRIVDFISRE 193
 126 CGYKATVADGKIQGPPKCAAGNOSTITCIDLYNNPQRQALRSAPAEFORLEVIATV 185
 194 AVHINWFSRKRIGANRADVHASTSSRLDAIRSYGASVVDLEIKVSYDADNDFE 253
 166 AVHNRVGFILRKQDQAPALRTFVASSRSSEINRIIGAAISKELEF-----SHDEYV 240
 254 KMDG--YISNANYAKTITMLIFINDRLVDCTALKRAIEFYVSNLTPQASKPEIYSHL 311
 241 KEAEDELITQVYSAKCOMLFTINQRIESTALRTSVDSIYATYLRGHHPYVMSLTL 300
 312 PPSHDVNIHPKKEVSLNOERIEETIRNAIEEKLMSNTTRIFQOALNSGIAQANP 371
 301 PPDNDVNHPTKHEVHLYGEIYDSIKOOVEALNLSNATRFYFQ--LRLPRA-----P 355
 372 QKDVSASMGSGTSQKTFPVQWRTDP-----RNPGRILTYWHGQSSNL 418
 356 DLEDTOLAD-----KTQRIYKEMWNTDSTEQKDKFLAPLYKSDSGVSSSSQSEASRLP 410
 419 EKKFDLAVSVNVRSSRNQKADGLSSRHELLIVEIDSSFPGLDIYKNCYVGLADEAF 478
 411 EESFRVIAAK---KSREVR-----LSSVDMKRIVERQCSVOLRSTLNLVYGCVDERR 462
 479 ALIQRNRLVYNNVNIISKELMVOALCFRGNFNAITLSEPAQLQELLYMALKDELMSD 538
 463 ALFQHEIRLYMCMTRSFSEELFYQRMIFRQNSLETISPLPLKELTILSEBAAGWT 522
 539 EKDEKELIAVNTTEILKENAMENEFYSHIDQKTLRPVYLDQYTPDDRLPEVYL 598
 533 PEDGKAEIADGADILLKKNAPIMREYFGIRISEDGLSELPSELHQHRCVAHLPVYLL 582
 599 ALGNDVTWDEKECPRIYASVGNFALHPPLIPNPSGNGIHLYKKNRDSMADEHAENDL 658
 583 RLATFVDEWEOETRCETFCRETARFY----- 608
 659 ISDENVDVQELIAEAAAMAO-----REWTOHYVLEPSPMRPLFKPKSMATOGTFVQVA 712
 609 -----AQIDMREGATAGFSRWIMHEVLEPFAEKYLLPPRIKDO--IYEEL 652
 713 SLEKTYIERC 724
 653 NLEFLYKVERC 664

RESULT 14
 AAB85849 standard; protein: 769 AA.
 ID AAB85849
 AC AAB85849;
 XX 29-OCT-2001 (first entry)
 DT
 DE Yeast MHL1 protein.
 XX
 KW Hypermutable bacteria; mismatch repair gene; MMR gene; Muts; MutL;
 RV MutY; PMS2; MHL1; MHL3; PMSR; biocatalysis; bioremediation; biochemical;
 FW drug discovery; detoxification; toxin; biotransformation.
 XX
 OS Saccharomyces cerevisiae.
 XX
 PN WO200159092-A2.
 XX 16-AUG-2001.
 PD
 XX 12-FEB-2001; 2001WO-US04339.
 PF
 XX 11-FEB-2000; 2000US-0181929.
 PR
 XX (UYJO) UNIV JOHNS HOPKINS.
 PA
 XX Nicolaides NC, Sass PM, Grasso L, Vogelstein B, Kinzler KW,
 PI

XX WP1: 2001-514664/56.
 DR N-ESDB; AAH76362.
 XX Making hypermutable bacteria for biocatalysis, bioremediation and drug
 PT discovery, involves introducing polynucleotide comprising dominant
 PT negative allele of mismatch repair gene under regulatory sequence
 PT control
 XX
 PS Example 1; Page 37-38; 66pp; English.
 XX
 XX The invention provides a method for generating a hypermutable bacteria.
 XX The method involves introducing a polynucleotide having a dominant
 XX negative allele of a mismatch repair (MMR) gene under the control of an
 XX inducible transcription regulatory sequence, into a bacterium. The cell
 XX becomes inducibly hypermutable. The method is useful to create desirable
 XX output traits for commercial applications, using dominant negative
 XX alleles of mismatch repair proteins. The mismatch repair gene is a Muth,
 XX Muts, MutL or MSH homologues. The hypermutable bacteria is useful for the
 XX PMSR or PMSR homologues. The hypermutable bacteria is useful for the
 XX production, biocatalysis, bioremediation and drug discovery. It is also
 XX useful in manufacturing industry for the generation of new biochemicals
 XX useful for detoxifying noxious chemicals from by-products of
 XX manufacturing processes or those used as catalysts, for remediation of
 XX toxins present in the environment including polychlorobenzene, heavy
 XX metals and other environmental hazards for which there is a need to
 XX remove them from the environment. The hypermutable bacteria is also
 XX useful for screening novel mutations in a gene or a set of genes that
 XX produce variant siblings that exhibit a new output trait not found in
 XX wild type cells. The bacteria are also useful for producing increased
 XX quantity or quality of protein or non-protein therapeutic molecule e.g.
 XX penicillin G, erythromycin and clavulanic acid, by biotransformation.
 XX Dominant negative alleles of the MMR gene are useful for producing higher
 XX quantities of recombinant polypeptides. The present sequence represents
 XX a yeast MHL1 protein.
 XX
 SQ Sequence 769 AA;
 29.7%; Score 1101; DB 22; Length 769;
 Query: Match 32.8%; Pred. No. 6,8e-82;
 Basic local similarity 157; Mismatches 243; Indels 144; Gaps 19;
 Matches 265; Conservative 157; Mismatches 243; Indels 144; Gaps 19;
 17 RIRLEESVNRIRIAGEVYORPSSAVKELTENSIGASSVAVVADGKLQVSDDG 76
 4 RIKALDASVNVKIAAGELLISPVNALEKEMNSIDANATMIDLYVEGGIKVQLTDNCS 63
 77 GIREFDLALICERHTSKLSAYEDLOTIKSGFERGALASMTYVGYVYTTTREGOLHEV 136
 64 GINKADLPICERFTTSLQKFEDLSQIOTYGFGEALASISHVARTVTKYKEDRCAM 123
 137 RVSYRDGVNENEPKCAAVKGTQVWENLFYNNVARKTLQNSNDYPRIVDFISRAVAH 196
 124 RVSYAGKMLSPKPVAGKGTTLVEDELFNIRSLRALRSHNDEYSKILDVVGRYAIH 183
 197 HINWFSRKRIGANRADVHASTSSRLDAIRSYGASVVDLEIKVS-YEDADSFEM 255
 184 SKDIGFCKKFGDSNLSVKPSYVQDRIRTFENKSVASMLITFHLISKEDL--NLESV 241
 256 DGYISNANYAKK-TIMLIFINDRLVDCTALKRAIEFYVSNLTPQASKPEIYSHL 314
 242 DGKVCNLFNSKSIISLIFINNRLVTCDLRLANYSNTLPGRFPYILGIVIDPA 301
 315 HDVNIHPKKEVSLNOERIEETIRNAIEEKLMSNTTRIFQOALNSGIAQANPKD 374
 302 AVDVNHPTKHEVHLYGEIYDSIKOOVEALNLSNATRFYFQ--LRLPRA-----P 355
 375 KV-----SEASMGSGTSQKTFPVQWRTDP--RNPGRILTYWHGQSSNL 411
 362 TIESDRNKRSLRQAVVENSYTTANSQLRAKQKQENKVLVRI--SQAKITFSLSSQOF 419
 412 HGQSSNLEKKFDLAVSVNVRSSRNQK-----DA-----GD- 442
 420 NFGSSTRKQLSEKRVYVNSHQAELTLNESQPDANTINDNDLKDPKKKQKGLGY 479

QY 443 -----LSSRHELVLEIDSSPHGLLDIVKNCYV 471
 DB 480 KVSIADEKNAIPISMDGYIRVPEKERVNVLISIKKLRKVDSDIHRRELDIFANLNV 539
 QY 472 GLADE--AFALIOHNRILYVNVNISKELMAYOALCRFENFNAIQSEAPLOELLVMA 529
 DB 540 GYVDEERRRLAIOHDKLFLIDIGSVCEYELFYQIGLDFANFKINQSTNVSDIYLVN 599
 QY 530 L--KDELMSEKDEKLEIAEVTTELKENAMINEYFSIHIDOG-----KLTRL 579
 DB 600 LLSFDEL--NDASKERI--ISKI-----WDMSSMLNEYISIELVNDGLDNDLKVLSL 652
 QY 580 PVLVDQYTPMDRLPEFVIALGNDVTWDEKECFRTVASAGNFYALHPILPNPSCNGI 639
 DB 653 PLLKGYIPSLVLPFIYIRLGEVDEWDEQECIDGILREIALLY-----IP----- 699
 QY 640 HLYKKNDSMADEHAENDLISDENDVOELLAEAMAREMTI-----OHVLEPSPRLF 695
 DB 700 -----DMVPKVDYTLDAISLSEDEKAFINRKEHISLSLEHVLFPCKIKR 742
 QY 696 LKPPKSMATDGTFOVASLEKLYIFERC 724
 DB 743 FLAPRHILND--VEELANLPDLIKVEERC 769

RESULT 15

AAG63951
 ID AAG63951 standard; protein; 769 AA.

AC AAG63951;

DT 29-OCT-2001 (first entry)

DE Amino acid sequence of yeast mismatch repair protein MLH1.

XX MLH1, mismatch repair gene; MMR gene; hypermutable yeast.

OS Unidentified.

XX MO200162945-A1.

PD 30-ANG-2001.

XX 21-FEB-2001; 2001MO-US05447.

XX 23-FEB-2000; 2000US-0184336.

XX (UYJO) UNIV JOHNS HOPKINS.

PA (NICO/) NICOLAIDES N C.

PA (SASS/) SASS P M.

PA (GRAS/) GRASSO L.

PA (VOGE/) VOGELSTEIN B.

PA (KINZ/) KINZLER K W.

PI Nicolaides NC, Sassi PM, Grasso L, Vogelstein B, Kinzler KW;

DR WPI; 2001-522820/57.

DR N-PSDB; AAH75039.

PT Making hypermutable yeast that exhibit novel selected output traits for

PT commercial applications, comprises introducing polynucleotide

PT containing dominant negative allele of mismatch repair gene

XX Disclosure; Page 36; 60pp; English.

XX The present sequence represents yeast MLH1. MLH1 is a mismatch repair

CC (MMR) gene. The specification describes a method for making a

CC hypermutable yeast, comprising introducing a polynucleotide containing

CC a dominant negative allele of a mismatch repair (MMR) gene, into a

CC yeast, whereby the cell becomes hypermutable. The method is useful

CC to create desirable output traits for commercial applications, using

CC dominant negative alleles of mismatch repair proteins. The hypermutable

CC yeast is useful for production, biocatalysis, bioremediation and drug
 CC discovery. It is also useful in genetic screens for the direct selection
 CC of variant subclones that exhibit new output traits. The hypermutable
 CC yeast is also useful in the manufacturing industry for the generation
 CC of new biochemicals, for detoxifying noxious chemicals from by-products
 CC of manufacturing processes or those used as catalysts, for remediation
 CC of toxins present in the environment including polychlorobenzenes, heavy
 CC metals and other environmental hazards for which there is a need to
 CC remove them from the environment. The yeast is also useful for producing
 CC increased quantity or quality of protein or non-protein therapeutic
 CC molecule e.g., Penicillin G, Erythromycin and Clavulanic acid, by
 CC biotransformation.

XX Sequence 769 AA;

Query Match 29.7%; Score 1101; DB 22; Length 769;
 Best Local Similarity 32.8%; Pred. No. 6 8e-82;
 Matches 265; Conservative 157; Mismatches 243; Indels 144; Gaps 19;

QY 17 RIRLEESVNRIRAGEVIGRPSAVKELIENSLDAGSSVAVKDGKLIQVSDGH 76
 DB 4 RIKALDASVYVNTKIAAGEIILISPVNALEKEMENSIDANATMIDILYKGGIKVQITDNGS 63
 QY 77 GIREDELAIICERHTTSKLSAYEDLOTIKSMGERGELASMTYGVHTVTITTEGOLHY 136
 DB 64 GIMKADIPILCERHTTSKLOKFEELSQIQTYGFEGLASISHVAVYTTTKVEDRCAN 123
 QY 137 RVGYROGVMEPEKPCAAVAGCTQVMVENLFFYNVARKKTIQNSDDQPKIVDFISRAVH 196
 DB 124 RVSTAEKMLSEPPKPAVGKDGTTILVEDLDFNIPRLRALRSHDESKILDVVGRFAH 183
 QY 197 HINVFSCRKRGANRADVHASSTSRDLATRSYVAGSVBDELEIKVS--YEDAADSIFFKM 255
 DB 184 SKDIGCFCKKFGDSNVSLSVAPSTYQDRIRTYFNKSVANLITPHISKVEDL--NLESV 241
 QY 256 DGYISNANYAKK--TIMILINDRLVDCALRAIEFVYSATLPQASKPTIYSIHLPSE 314
 DB 242 DGKVCNINFTSKSISILFIENRLVYCDLLRRLANSVSNYLPKGRFPFIYIGIYDPA 301
 QY 315 HVDVNIHPTKREYSLNOERIEETIRNAIEBKLNSTNTRIFQOVALNGIAQANPKD 374
 DB 302 AVDVNHPTRKREVFLSODELIEKIANQIAELSAIDTSRTFKASSISTPKPESLIPND 361
 QY 375 KV-----SEASMGSGTKSOKIPVSOVTRDPNRPGRHTY-----W 411
 DB 362 TIESDRNRKSLRQAVYENSYTTANSQLRAKQENKILVRIDA--SQAKITPSFLSSQOF 419
 QY 412 HGSSNLEKKFDLYSVNVYRSRRNOK-----DA-----GD--442
 DB 420 NFEGSSYTRQLSEPKYVNVSHQAEKTLNSESOPDANTINDNDLKDPKKKQKLGDY 479
 QY 443 -----LSSRHELVLEIDSSPHGLLDIVKNCYV 471
 DB 480 KVSIADEKNAIPISMDGYIRVPEKERVNVLISIKKLRKVDSDIHRRELDIFANLNV 539
 QY 472 GLADE--AFALIOHNRILYVNVNISKELMAYOALCRFENFNAIQSEAPLOELLVMA 529
 DB 540 GYVDEERRRLAIOHDKLFLIDIGSVCEYELFYQIGLDFANFKINQSTNVSDIYLVN 599
 QY 530 L--KDELMSEKDEKLEIAEVTTELKENAMINEYFSIHIDOG-----KLTRL 579
 DB 600 LLSFDEL--NDASKERI--ISKI-----WDMSSMLNEYISIELVNDGLDNDLKVLSL 652
 QY 580 PVLVDQYTPMDRLPEFVIALGNDVTWDEKECFRTVASAGNFYALHPILPNPSCNGI 639
 DB 653 PLLKGYIPSLVLPFIYIRLGEVDEWDEQECIDGILREIALLY-----IP----- 699
 QY 640 HLYKKNDSMADEHAENDLISDENDVOELLAEAMAREMTI-----OHVLEPSPRLF 695
 DB 700 -----DMVPKVDYTLDAISLSEDEKAFINRKEHISLSLEHVLFPCKIKR 742
 QY 696 LKPPKSMATDGTFOVASLEKLYIFERC 724

Mon Apr 7 09:23:53 2003

DB 743 FIAPRHUKD--VEINLMDLKVFERC 769

Search completed: March 27, 2003, 15:43:57
JOB time : 44 secs

us-09-954-950-2.rag


```

Db 193 DYKRIYDLISMAIHHNVNSVSCRRHGAVKADVHSVSPSLDSIRSVYGVSAVKMLKV 252
Oy 242 KVSIEDAADSIFFKADGYISANNAYAKKTTMLFINDRLVDCTALKRAIEFYSATLPAS 301
Db 253 EVSSDSSGCGTFDEGEFISNSNYAKKTIILVFINDRIVECSALKRAIEIYATLPAS 312
Oy 302 KPEIYMSHLESEHVDVNIHPTKKEVSLINOERIETIRNAIEEKLMNSNTTRIFOTAL 361
Db 313 KPEIYMSHLESEHVDVNIHPTKKEVSLINOERIETIRNAIEEKLMNSNTTRIFOTAL 372
Oy 362 NLSGIAQANPKDKVSEASMGSGTKSQKIPVSQWVRIDPDRNPSCGRLHTYWHGQSSNLEKK 421
Db 373 EYIQ-STLSQKSDSPVQSKPSCGKTQCVPKWRTSDSPAGRLHAFLOPKPSLDPK 431
Oy 422 FDLVS-VANNVSRNOKDADGLSSRHLLVEIDSSFPGLDIDYKNTCTYVGLADEARAL 480
Db 432 VSSLSVRSVVRQRNPKEADLSSVQELIAGVDSCHPGLMETVYRNCTYVGMADVPAL 491
Oy 481 IOHNRRLVNVVNIISKELTQOALCRFGNFNAIQISEPAPLOELIYVALKDEDEL--MSD 538
Db 492 VOYNHLHLVNVNLSKELTQOALCRFGNFNAIQISEPAPLOELIYVALKDEDELPGND 551
Oy 539 EKDEKLEIAEVNTEILKENEMINEYFSIHIDOGKLTRLPVVLDQYTPMDRLPEVL 598
Db 552 TKDLEKLEIAEVNTEILKENEMINEYFSIHIDOGKLTRLPVVLDQYTPMDRVPEVL 611
Oy 599 ALGNVYVNDDEKCECFRTYASAVNGFYALHPILPNSGNGIHLKYKNDSADHAENDL 658
Db 612 CLGNVVEDEKCECFRTYASAVNGFYALHPILPNSGNGIHLKYKNDSADHAENDL 671
Oy 659 ISDENVDQOELIAEAAMAOREWITQHVLFPSMRLFLKPKSMATDGTFOVASLEKLY 718
Db 672 VDMENNDQDILSDAENMAOQREMSIQHVLPSPMLFLKPKSMASNGFYVVASLEKLY 731
Oy 719 KIFERC 724
Db 732 KIFERC 737

```

RESULT 2
US-10-079-429-2
Sequence 2, Application US/10079429
Publication No. US20030027177A1
GENERAL INFORMATION:
APPLICANT: Haseltine et al.
TITLE OF INVENTION: Human DNA Mismatch Repair Proteins
FILE REFERENCE: PFI06P3D1
CURRENT FILING DATE: 2002-02-22
PRIOR APPLICATION NUMBER: PCT/US95/01035
PRIOR FILING DATE: 1995-01-25
PRIOR APPLICATION NUMBER: 08/466,024
PRIOR FILING DATE: 1995-06-06
PRIOR APPLICATION NUMBER: 08/465,769
PRIOR FILING DATE: 1995-06-06
PRIOR APPLICATION NUMBER: 08/294,312
PRIOR FILING DATE: 1994-08-23
PRIOR APPLICATION NUMBER: 08/210,143
PRIOR FILING DATE: 1994-03-16
PRIOR APPLICATION NUMBER: 08/187,757
PRIOR FILING DATE: 1994-01-27
NUMBER OF SEQ ID NOS: 78
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 756
TYPE: PRT
ORGANISM: homo sapiens
US-10-079-429-2

Query Match 39.2%, Score 1453, DB 9, Length 756;
Best Local Similarity 39.7%, Pred. No. 2, 7e-104;
Matches 312; Conservative 144; Mismatches 214; Indels 116; Gaps 11;

```

Oy 18 IRRLEESVNRILAGEYIORPSSAVKELIENSIDAGASSVAVKDGLKLIQVSDG 77
Db 8 IRRLEEVNRILAGEYIORPSSAVKELIENSIDAGASSVAVKDGLKLIQVSDG 67
Oy 78 IRFEDLALICERHTTSKLSAYEDLOTKSMGFRGELASMTYGHVTVTTTSQHLGR 137
Db 68 IRKEDLALICERHTTSKLSAYEDLOTKSMGFRGELASMTYGHVTVTTTSQHLGR 127
Oy 138 VSYRSDGVNENPKCAVNGTQVAVENLFYMMVARKKTIONSDYKIDVFISPAVH 197
Db 128 ASYSDGKLKAPKPCAGNQGQITVEDLFTYIATRRALKPKSEYKILIEVGRYVH 187
Oy 198 INVFSCRKHGANPDYHASSTSSRLDAIRSVYASVRLIEIKVSEYEDAADSIFFKDG 257
Db 188 AGISEFYAKKGEIYADVADRTLPNASTVDNIRSVFENAVSRELIIEICDCKTLA--FMMG 244
Oy 258 YISNANVYAKKTIILFINDRLVDCTALKRAIEFYSATLPQASKPTIYMSHLPSEHVD 317
Db 245 YISNANVSVKCIPLFLFNHRLVESTSLRKAIEFYVAAYLPKNTHPPLYLISLEISPNVD 304
Oy 318 VNIHPTKKEVSLINOERIETIRNAIEEKLMNSNTTRIFOTALNLSGIAQANQ--KDK 375
Db 305 VNIHPTKKEVSLINOERIETIRNAIEEKLMNSNTTRIFOTALNLSGIAQANQ--KDK 363
Oy 376 VSEASMGSGTKSQKIPVSQWVRIDPDRNPSCGRLHTYWHGQSSNLEKK 426
Db 364 TSLSSSTSSGSDKVAQWVRTDSRQ--KLDAFIQPLSKPLSSQPAIYTEKTDISS 421
Oy 427 VRNVYR-----SRNOKDAGLS----- 444
Db 422 GRARODEMELDAPAEVAAKNQSLBDTTKGTSSEMSKRGPTSSNPKRRHREDSDVM 481
Oy 445 -----SRHELIV-----RIDSSFHFGLDIYKNCYVGLADEAPALI 481
Db 482 VEDSRKRENTACIPRRRIINLTSLVLSQEBINDEGHVLAEMLNHSFVCVAPQNALA 541
Oy 542 OHOFKLYLNTTKLSEELFYQILLYDFANFGVLRISERAPFLDLAMIALDPSGWTED 601
Db 542 DEKLEIAEVNTEILKENEMINEYFSIHIDOGKLTRLPVVLDQYTPMDRLPEVL 601
Oy 602 GPKREGIAIYEFLEKKKEMADYFSLIDEIGNILGPLIDNVPLEBPLFILELA 661
Db 602 NDVYMDDEKCECFRTYASAVNGFYALHPILPNSGNGIHLKYKNDSADHAENDLSD 661
Oy 662 TEVNDDEKCECFRTYASAVNGFYALHPILPNSGNGIHLKYKNDSADHAENDLSD 693
Db 662 ENDV---DOELIAEAAMAOREWITQHVLFPSMRLFLKPKSMATDGTFOVASLEKLY 718
Oy 694 ESTLSGQSEVAGSIPNSM---KMTVEIYVKAIRSHILPKHFTEDGNITLQNLNPLLY 750
Db 719 KIFERC 724
Db 751 KIFERC 756

```

RESULT 3
US-08-788-657-20
Sequence 20, Application US/09788657
Patent No. US20020123149A1
GENERAL INFORMATION:
APPLICANT: Sasse, Phillip
APPLICANT: Kinzler, Kenneth
APPLICANT: Grasso, Luigi
APPLICANT: Vogelstein, Bert
TITLE OF INVENTION: Methods for generating hypermutable
FILE REFERENCE: 01107.00097
CURRENT FILING DATE: 2001-02-21

PRIOR APPLICATION NUMBER: 60/184,336
PRIOR FILING DATE: 2000-02-23
NUMBER OF SEQ ID NOS: 25
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 20
LENGTH: 756
TYPE: prt
ORGANISM: Homo sapiens
US-09-788-657-20

Query Match 39.1%; Score 1452; DB 10; Length 756;
Best Local Similarity 39.6%; Pred. No. 3,2e-104;
Matches 311; Conservative 145; Mismatches 214; Indels 116; Gaps 11;

18 IRRLEESVNRNIAAGVIOIPSSAVKELIENSIDAGASSVAVKGGKLIKIOVSDGCH 77
18 IRRLEESVNRNIAAGVIOIPSSAVKELIENSIDAGASSVAVKGGKLIKIOVSDGCH 77
8 IRRLEESVNRNIAAGVIOIPSSAVKELIENSIDAGASSVAVKGGKLIKIOVSDGCH 67
78 IRRLEESVNRNIAAGVIOIPSSAVKELIENSIDAGASSVAVKGGKLIKIOVSDGCH 137
68 IRRLEESVNRNIAAGVIOIPSSAVKELIENSIDAGASSVAVKGGKLIKIOVSDGCH 127
138 VSYRQGVMEKNEPPCAAVAGTQVMEVNFYNNVARRKKTIONSNDYPKIVDFISRAVH 197
128 ASYSDGKLAAPKPCAGNGOTQVMEVNFYNNVARRKKTIONSNDYPKIVDFISRAVH 187
198 INVETSCRRHGANRADVHASTSSRLDAIRSYVAGSVADLEIKVSEDAADSIKFMG 257
188 AGISFVKKOGETVADVRLPNASTVDNIRISGNVSRLEIIGEDKLA---FKMG 244
258 YISNNAVYAKKTIIMIFINDRLVDCATKRAIEFYSAIIPQASKPTIYSIHLPSEHVD 317
245 YISNNAVYAKKTIIMIFINDRLVDCATKRAIEFYSAIIPQASKPTIYSIHLPSEHVD 304
318 VNIHPKKEVSLNOERIIETIRNAIEKLNSTTRIFOTQALNLSGIAQAMPD---KDK 375
305 VANHPKHEVHFLHESILIERVOQHIESKILGSSRMATYOTL-LPGLAGPBGEMVKT 363
376 VSEASMGSTKSOIKIPVSGWRTDPRNPSGRILTYHNGOSSNL-----EKKEDLVAS 426
364 TSITSSSTGSSSDKYVAHQMVRTDSREQ--KIDAFIQLPSKPLSQPOATVETDKTDISS 421
427 VRRVVR-----SRNOKDAGDLS-----444
422 GRARODEEMLELPADAEVAAKNQSLGDTTKGTSEMSEKRGPTSSNPKRHREDSDEV 481
445 -----SRNELLV-----EIDSSPHGLDLYVKNCTYGLADEAPALI 481
482 VEDDSKKEMTACTPRRIINLTSTVLSLOEINQGHVLEKMLHNSFVGCYVPMQWALA 541
482 QHNTRLYLNVNVIISKELMYOALCFRGNFNAIOLSEPAILOELVMAKDELMSEDK 541
542 QHNTRLYLNVNVIISKELMYOALCFRGNFNAIOLSEPAILOELVMAKDELMSEDK 601
542 DELELAENVTELKEMAKINIEFISIHDOCKLRLPYVLDQYTPDDRLPEVALG 601
602 GPEGLAEYIVELKRAKAMLADEYFLEIDEENLIGLPLLDNYYVPPLEGDIFILIRLA 661
602 NDVTWDEKCEFTVASAVGNFALHPPLIPNSGNGIHYKKNRSMADENHENDLIS 661
662 TEVNMDEKCEFTVASAVGNFALHPPLIPNSGNGIHYKKNRSMADENHENDLIS 661
662 ENDV---IDELLAEAANAQREMTIQHVLPSMRLELPRKPMADGTFVQVASELEK 718
694 ESTLSQOSEVPSIPNSW---KMTVEHLYKALRSHILPRKHFTEGNTILOANLPDLV 750
719 KIFERC 724
751 KIFERC 756

RESULT 4
US-09-788-657-15

Sequence 15, Application US/09788657
Patent No. US20020123149A1
GENERAL INFORMATION:
APPLICANT: Nicolaides, Nicholas
APPLICANT: Sassi, Philip
APPLICANT: Kinzler, Kenneth
APPLICANT: Grasso, Luigi
APPLICANT: Vogelstein, Bert
TITLE OF INVENTION: Methods for generating hypermutable
FILE REFERENCE: 01107.00097
CURRENT APPLICATION NUMBER: US/09788,657
PRIOR FILING DATE: 2001-02-21
PRIOR APPLICATION NUMBER: 60/184,336
PRIOR FILING DATE: 2000-02-23
NUMBER OF SEQ ID NOS: 25
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 15
LENGTH: 769
TYPE: prt
ORGANISM: Saccharomyces cerevisiae
US-09-788-657-15

Query Match 29.7%; Score 1101; DB 10; Length 769;
Best Local Similarity 32.8%; Pred. No. 5,2e-77;
Matches 265; Conservative 157; Mismatches 243; Indels 144; Gaps 19;

17 IRRLEESVNRNIAAGVIOIPSSAVKELIENSIDAGASSVAVKGGKLIKIOVSDGCH 76
4 RIKALDSVYVNRNIAAGVIOIPSSAVKELIENSIDAGASSVAVKGGKLIKIOVSDGCH 63
77 GIREFDELAICERHTTSKLSAYEDLOTIKSMGREGALASMTYVGHVTTTTEGOLHGY 136
64 GINKADPLICERHTTSKLSAYEDLOTIKSMGREGALASMTYVGHVTTTTEGOLHGY 123
137 RVSYRQGVMEKNEPPCAAVAGTQVMEVNFYNNVARRKKTIONSNDYPKIVDFISRAVH 196
124 RVSYRQGVMEKNEPPCAAVAGTQVMEVNFYNNVARRKKTIONSNDYPKIVDFISRAVH 183
197 HINVTSCRRHGANRADVHASTSSRLDAIRSYVAGSVADLEIKVSEDAADSIKFM 255
184 SKDIGSCKFFGSSNTSLVSKPSTVODRIRTVFNNSVANSNLTFFHISKVED--NLESV 241
256 DGYISNNAVYAKK-TIMIFINDRLVDCATKRAIEFYSAIIPQASKPTIYSIHLPSE 314
242 DGYISNNAVYAKK-TIMIFINDRLVDCATKRAIEFYSAIIPQASKPTIYSIHLPSE 301
315 HDVNIHPKKEVSLNOERIIETIRNAIEKLNSTTRIFOTQALNLSGIAQAMPD 374
302 AVDVNHPKKEVSLNOERIIETIRNAIEKLNSTTRIFOTQALNLSGIAQAMPD 361
375 KV-----SEASMGSTKSOIKIPVSGWRTDPRNPSGRILTY-----W 411
362 TIESDRNRKSLROAQVENSITANSOLRAKROENLVIDA--SOAKITPSSSQOF 419
412 HGSSNLEKEDLVSVNVRNRRNOK-----DA-----GD- 442
420 NFEGSSSTRQLSPKVTYNVSHQEAELTLNESQPDATNTINDNLDKQPKKKQKGDY 479
443 -----LSRHELVLEIDSSPHGLDLYVKNCTYV 471
480 KVSIADEKRNALPISKGYIRVPERVNVNLTISIKRLRKVDSDSIRELTDIFANLNV 539
472 GLADE--AFALIOHNPRLYLNVNVIISKELMYOALCFRGNFNAIOLSEPAILOELVMA 529
540 GYVDEERRLAIOHDLKFLIDYGSVCYELFYQIGLDFRNFKINLOSTNVSDDYLYN 599
530 L--KDELMSEDEKDEKLEIAENVTELKEMAKINIEFISIHDOCKLRLPYVLDQYTPDDRLPEVALG 579
600 LSEFDEL--NDASKERK--ISKI-----WDMSSMLNIEYSIELVNOGLDNDLKSVKLSL 652
580 PVVLDQYTPDDRLPEVALGNDVYTWDEKCEFTVASAVGNFALHPPLIPNSGNGI 639

Db 653 PULLKTIISLVLPFFITRLGKEVMEDEQELDGLIREALLY-----IP----- 699
QY 640 HLYKKNRDSMADEHNDLISDENVDQELIAEAAAMOREWTI-----OHVFPSPMRLE 695
Db 700 -----DWPVKYDTLDASISEDEKQFIRKHHISLLEHVLPFCIKRR 742
QY 696 LKPKSMADGTGVQVASELKYIFERC 724
Db 743 FLAPRHILD--VEEIANLPDLKYFERC 769

RESULT 5
US-09-815-242-5262
Sequence 5262, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815.242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5262
LENGTH: 669
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-5262

Query Match 16.1%, Score 599, DB 10, Length 669;
Best Local Similarity 27.2%, Pred. No. 3.4e-38;
Matches 179; Conservative 115; Mismatches 203; Indels 160; Gaps 19;

QY 17 RIRLEESVNRVIAAGVIOIPSSAVKELIENSIDAGASSVAVKDGKLIQVSDGCH 76
Db 3 KIKELQTLAKIAGVVERPSSVVKLLNADAGTEISIEVEESGVOSIRVWDGS 62
QY 77 GIREEDALICERHTTSLASVEDLOTIKSGFGEALASTYGVHVVTITTEGOLHG 136
Db 63 GIEADGLVPHRATSLDDEDELPHRTIGFGEALASISSAVKYLKCTD--NANGN 121
QY 137 RVSRTDGMVEPPKCAVKTGVAVENTFTNMVARKKTLONSDDYPKIVDFISRPVH 196
Db 122 EIVYENGELINH--KPAKAKCTDILVESLFFYTPARLKIKYKSLVTELGKIDIVNRAMS 180
QY 197 HINTFSCRKGANRADVHASTSSRLDAISVYGVASVNRRLIILKSYENAAISIRMD 256
Db 181 HPDIRIALISGKTMSTNGSGRTN--EVMETIGMVARVAHIS--GPTSD--YHIE 233
QY 257 GYI-----SNANYVAKKITMIFINDRLVDCTALKRAIFVYSATLPQSKRFIYMSI 309

Db 234 GFVAKPEHSRKNKHYIS-----IFNGRYIKNEKMLAKILLEGYHLLTIGRPICYYNI 287
QY 310 HLPSEVVDVNIHPTKEVSLVNOERITETIRNAIEKILMNSNTTRIFOTAL-----NLSG 365
Db 288 EMDPLIVDVNVHPTKLEVLISKEEOIYQLIVSKIOE-----AFKDRILIPKNNIDY 338
QY 366 IAOANPOKDVSEASWGSGTSOKIIPVSQMVRTDRNPSGRILTYWHQSSNLEKPPDLV 425
Db 339 V-----PKKKVLIHS-----FEOKIEFEORONTE-----NNOEKTFSSSE 373
QY 426 SVRNVRYSRRNCKD-----AGDLS-----RHEL 449
Db 374 ESNKRPFMAENONDEIVIKEDSYNPFVTKTSLLADDESSYNNMRKDEDFPKKOEI 433
QY 450 LYEIDSSPHGLDLYKN-----CTYGLADEAPAL 480
Db 434 LQEMQTFSDNDTSVQNTENKASDDYDVNDIKGTSKDPKRRIPYMEIYQVHGTYII 493
QY 481 IOHNTRLVNVVNISSKELMYQALCRFNFNAIQLSBPAPLOELLVMAI-----KDEELM 536
Db 494 AQNEFGMTMIDQHAQERIKYTFYDKIGEVY-----NEVODLLPLTFHFSKDEQLV 546
QY 537 SPEKDEKLEIA-----EVTETILKENAVEMINEYFSIHI 570
Db 547 IDQYKNEIQVGMLEHFGHDYIVSYVPWPKDEVE--EIKDMIELILEKKYDI 602

RESULT 6
US-09-815-242-12334
Sequence 12334, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815.242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12334
LENGTH: 669
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12334

Query Match 16.1%, Score 599, DB 10, Length 669;
Best Local Similarity 27.2%, Pred. No. 3.4e-38;
Matches 179; Conservative 115; Mismatches 203; Indels 160; Gaps 19;

QY 17 RIRLEESVNRVIAAGVIOIPSSAVKELIENSIDAGASSVAVKDGKLIQVSDGCH 76

PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO.13206
 LENGTH: 649
 TYPE: PRT
 ORGANISM: Streptococcus pneumoniae
 US-09-815-242-13206

Query Match 15.2%; Score 565; DB 10; Length 649;
 Best Local Similarity 27.7%; Pred. No.1,4e-35;
 Matches 162; Conservative 115; Mismatches 219; Indels 88; Gaps 16

18 IRRLEESVNRIRAGEVIQRPSSAAVKELIENSLDAGASSVAAYKDGIKLQVSDDG 77
 Db 4 IIELEMLNIAAGEVIERPASYVKELEVENADIGSSQIIIEIEPGLKVKQITDNGH 63

78 IRFDLALCEHTTSKLSAYEDLOTFSKMGREGELASMTYVGHVYTTTIEGQLHGR 137
 Db 64 IAHDELVALRHRHASKIRKNQADLRIRIFLGRGALPLSIASVYTLTLTAADGASHGR 123

138 VSYRGVNENERPCPAAYKQIVQENLEPNYVARKKTQNSNDYPKLVDFSPRAVHH 197
 Db 124 LVARKGEVE-ETIPATSPVGTVCYCEDLEFNTPAKTKKSOALSHIIDVNRGLAH 182

198 INVFESCRKHGARNADVSHASTSRIDAIRSYGASVVDLIEIKSYEDPADSIFPKMG 257
 Db 183 PEISFSLSDG--KMTPTRAGTGQLRQALAGIYGLVSAKKMEIENS-----DLDEISG 235

258 YIS-----NANYAKKITMLTFINDRLVDTALKRAIEFYSATLPQASKPEIYMSIH 310
 Db 236 FVSLPELTPRANKNYIS-----LFTNGRIYKKNFLNRALIDFGSKLWGRFLAVIHH 289

311 LPSEHVDVNIHPTEKVESLNLQERII-----ETIRNAIE-----KLMNSNTR 354
 Db 290 IDPYADLVAVHEPTKQEVRSRKEKELMTLVSEAIANSLSKQTLIPDALENIASVYNNRER 349

355 IFQOT-----QALNLSGIAQANPQKDKVSEASMGSGTSOKIIPVSQMVYTRDPRNPGRH 408
 Db 350 VEQTLPLKENTLYYEKTEPRSPQSEVADYQVELDDEQDILTFPAKELLDRITRAKAKH 409

409 TYWHGSSNLEK---KEDLVSVRRVNSRRNQKADGLSSRHELLVEIDSSFHPGLDI 464
 Db 410 -FAERKPANYDQLDHPEDLASI-----DKAYDKLEREASFPFL--- 449

465 VKNCITYGLADEAPFLIQHNTFLYLVNVYNISKELMTYQALCRFGFNALQISEPADI 524
 Db 450 ---EFFQGMGTFLPAGRGDGLYTIIDHQAADERKYEEYRESIGVNDOSQ-----QQ 498

525 LLVMAI---KDELMSDEK---DDEKLEIAEV--NTEILKEN 558
 Db 499 LLVPIFFEPADDAIRLKERMPLEBVGVLAEYGENQFILNHH 542

RESULT 8
 US-09-815-242-13667
 Sequence 13667, Application US/09815242
 Patent No. US20020061569A1
 GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Karl L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 TITLE OF INVENTION: Prokaryotes
 FILE REFERENCE: ELITRA.011A
 CURRENT APPLICATION NUMBER: US/09/815.242
 CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 13667
 LENGTH: 650
 TYPE: PRT
 ORGANISM: Streptococcus pneumoniae
 US-09-815-242-13667

Query Match 15.2%; Score 565; DB 10; Length 650;

Best Local Similarity 27.7%; Pred. No. 1.4e-35; Matches 162; Conservative 115; Mismatches 219; Indels 88; Gaps 16;

18 IRRLEESVYVNRVNIAGEVIGORPSSAVKELIENSIDAGASSVAVAKDGKILQVSDGCHG 77
 4 IIELEPMLANOIAAGEVIERPASVYKELIENVIDAGSSQIIEIEEAGKAKKQIINDNGHG 63
 78 IREFDLALCEHRTSKLSAYEDLOTISMGFRGALASMTYVGVYVTTTTEGOLHGR 137
 64 IAHDEVELLRHARHSKIRKNOADIRIRTLGRGALPSIASVYLTLLTAVDGCASHGK 123
 138 VSYRDGVNENPKPCAANKGVQVWENEFYNNVARKKTLONSNDYPRIVDFISPAVHH 197
 124 LVARGGEVE-EVIRATISVGTVCYEDLEFNTPARLTKMSQOALSHIIDVNLGLAH 182
 198 INWFSCKHGANRADVHASTSRDLAIRSVYASVVRDLIEIKVSYEDADSLFKMDG 257
 183 PEISFSLSDG--KEMPTAGGOLRQAIAGIYGLVSAKKMEIRNS-----DLDFEISG 235
 258 YIS-----NANYAKKITMILFINDLVCTALAKRAIEFYVSKTLPOASKPFYMSIH 310
 236 FVSLPELTKRANKNIS-----LFTNGRYINFLNRAILDGFSGKLMVGRPLAVIH 289
 311 LPSEHVDVNIHPTKEVSLNORERII--ETIRNAIEE-----KIMSNTR 354
 290 IDPYLADVNVHPTKOEVAIRISKEKELMTLVSEALNSLKEQLLIPALNLAKSTYRNRK 349
 355 IFQF-----QALNLSGIAQANPOKDKYSEASMGSGTKSQKIPVSQWVRTDPRNPSGLH 408
 350 VEQTLPLKENTLLYKTEPRSPQTEVADYQVELTDEGQDLTLFAKETLRLTKRPAKHL 409
 409 TYWGOSSNLEK---KFDLVSRVNVVSRNNOADGLSRHSHLVEIDSSFFHGGLDI 464
 410 -FAEKRPANVQDLHPDLASTI-----DAYDKLEEESSSPEL--- 449
 465 VKNCTVGLADEAFALIQHNTRLVLYNVVNISKELMTQOALCRGFENAIQISEPAPLOE 524
 450 -----EFGQMGHTLFAQGRDGLYIIDQHAQERKYEYESJESIGNVQSQ-----QQ 498
 525 LLYVAL---KDDLEMSERK---DDEKLETAEV--NTEIILKEN 558
 499 LLYVYIEFPADALRLKERMPILLEEVGVFLAEGENQFILREH 542

RESULT 9

US-09-815-242-10961

; Sequence 10961, Application US/09815242.

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Karl L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 TITLE OF INVENTION: Prokaryotes
 FILE REFERENCE: ELITRA.011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 PRIOR FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 10961
 LENGTH: 629
 TYPE: PRT
 ORGANISM: Haemophilus influenzae
 US-09-815-242-10961

Query Match 14.8%; Score 547.5; DB 10; Length 629;

Best Local Similarity 27.3%; Pred. No. 3e-34; Matches 174; Conservative 119; Mismatches 251; Indels 93; Gaps 19;

18 IRRLEESVYVNRVNIAGEVIGORPSSAVKELIENSIDAGASSVAVAKDGKILQVSDGCHG 77
 3 IKILSPOLANOIAAGEVIERPASVYKELIENSIDGANKIIDIDENGANLIRLRDNCGG 62
 78 IREFDLALCEHRTSKLSAYEDLOTISMGFRGALASMTYVGVYVTTTTEGOLHGR 137
 63 IPKEBLSIALARHATSKLADLDLEALISLGFREGALASISVSRLTSTFEQTEQWQ 122
 138 VSYRDGVNENPKPCAANKGVQVWENEFYNNVARKKTLONSNDYPRIVDFISPAVHH 197
 123 VYAGRDMEETIKPASHVGTVEVANKLFNTPARRKLRTDKTEFFSHIDEVIRIALTK 182
 198 INWFSCKHGANRADVHSA--STSSRLDAIRSVYASVVRDLIEIKVSYEDADSLFKMD 256
 183 FNTAFETLHNCKIIRYRPAFAINQOLKRVANIGDDPVKALAIENKHD-----LHLS 237
 257 GYISNANYVAKKITK--ILFINDLVCTALAKRAIEFYVSKTLPOASKPFYMSIHLPSEH 315
 238 GWATPMSRTRNDLSYCYINGRWVRKVISHAIRQAAQYLPDAYPAVYLFIDLPHD 297
 316 VDVNIHPTKEVSLNORERIIETIRNAIEEKLMSNTRTRIFQOALNLSGIAQANPOKDK 375
 298 VDVNVHPTKEVSRHQORLHDFIYEGISHALNNOEOINMHTDS-----AVENHEENT 351
 376 VSE-----ASMGSGT-----KSKQIPVSQWVRTDPRNPSGLRHTYWHGOSS 416
 352 VREQPNVNSIRPNATAGQNSFAQYHEKPOQNPSPHSNTPMFMNHNSTGRDYRSPAPS 411
 417 NLEKKFDLVSVRNVSRNNOADGLSSRH-----ELLVEIDSSFFHPLGLDIYK 466
 412 KTEORLYAEALERTLPPTA--OKDISDPAQONISDTAKIISPEI---TECSSHALSLIE 466
 467 NCTVYGLADEAFALIQHNTRLVLYNVVNISKELMTQOALCRGFENAIQISEPAPLOEL 526
 467 N-----RALILOONQODFFLSTLEKLR--IQMOLAL-----KOIQIEQQLLP-I 509

QY 527 VALKDELMDSK--DDEKLEIAEYNTLLEKNAEMINEFSIHIDQDKLT--RLPV 582
 Db 510 VFFLTESQOAWOQYSDSK---KIGFEFIEHQADL-----RLTLNVPSSA 552
 QY 583 LDQYTPMDRLPEFVLTALGNDYTWDDKCEGFRVASA 619
 Db 553 L-----RTONLQKCVAM---LIRDENSSSFLTALCA 581

RESULT 10

US-09-815-242-10895
 : Sequence 10895, Application US/09815242
 : Patent No. US20020061569A1
 : GENERAL INFORMATION:
 : APPLICANT: Haselbeck, Robert
 : APPLICANT: Ohlsen, Karl L.
 : APPLICANT: Zyskind, Judith W.
 : APPLICANT: Wall, Daniel
 : APPLICANT: Trawick, John D.
 : APPLICANT: Carr, Grant J.
 : APPLICANT: Yamamoto, Robert T.
 : APPLICANT: Xu, H. Howard
 : TITLE OF INVENTION: Identification of Essential Genes in
 : FILE OF INVENTION: Prokaryotes
 : FILE REFERENCE: ELITRA.011A
 : CURRENT FILING DATE: 2001-03-21
 : PRIOR FILING DATE: 2000-03-21
 : PRIOR FILING DATE: 2000-03-21
 : PRIOR FILING DATE: 2000-03-21
 : PRIOR FILING DATE: 2000-05-23
 : PRIOR FILING DATE: 2000-05-23
 : PRIOR FILING DATE: 2000-05-26
 : PRIOR FILING DATE: 2000-05-26
 : PRIOR FILING DATE: 2000-10-23
 : PRIOR FILING DATE: 2000-10-23
 : PRIOR FILING DATE: 2000-11-27
 : PRIOR FILING DATE: 2000-11-27
 : PRIOR FILING DATE: 2000-12-22
 : PRIOR FILING DATE: 2000-12-22
 : PRIOR FILING DATE: 2001-02-16
 : NUMBER OF SEQ ID NOS: 14110
 : SOFTWARE: FASTSEQ for Windows Version 4.0
 : SEQ ID NO 10895
 : LENGTH: 710
 : TYPE: PRF
 : ORGANISM: Enterococcus faecalis
 : US-09-815-242-10895

Query Match 14.5%; Score 536; DB 10; Length 710;
 Best Local Similarity 26.1%; Pred. No. 2, 8e-33;
 Matches 179; Conservative 113; Mismatches 243; Indels 152; Gaps 21;

QY 17 RIRLEESVNRRIAGEVIRQSSAVKEIENSIDAGASSVAVKDGKILQVSDGH 76
 Db 3 KIOELSEOLANOIAAGEVERPASVYKELVENALDAGSQIDIFIEAGKTIQIDNGE 62
 QY 77 GIREFDLALCERHTSKLSAYEDLOTISMGREGALASMTYGVHTYTTIEGOLHG 136
 Db 63 GAKPEVLNAFKRHATSKLHTRDLEIRSLGREGALPSIASVSMIYETATAEEEGS 122
 QY 137 RVSRYDGVNENEPKCAAVKGOVVENLFYNNVARKKTIONSNDYPKVDISRPVH 196
 Db 123 YILKKGAYE-ENRPAALAKKTKMYSNLFYNTPALKTYKTIOTELANGDIYNRLALS 181
 QY 197 HINWTFSCRKHGANRADVHSA---STSSRLD---AIRSYGASVVDLIEIKVYEDAA 249
 Db 182 HPEVAFRL-----VHDGHKMSGTGNDKQTIAGIYGISTAKMKMI-----EGK 227
 QY 250 DSIFKKNDGYIS-----NANTYAKKITMILFINDRLDCTAKRAIEFYYSITLQASK 302
 Db 228 DUDFTLGTIVSLPEVTIRASRNTLS-----TTINGRYIKNFALNKAIVAGYGSKLWGRP 281

QY 303 PEIYMSIHLPSEHVDVNIHPTKREVSLLNOERILETIRNAIEEKIMNSNTTRLFQTOALN 362
 Db 282 PIAVLEIEMOPLLYVDVNVHPTKQEVRLSKERDLMALIEQAIOEVL---SQEOLIPNADN 338
 QY 363 LSGIAQANPKOKVSEKSMGSGTKRSOKTIPYSQWRTDPRP-SGRLLTYHAGGSSNLEKK 421
 Db 339 LR-FKKRLPEPKAEQOEIPLST--EQPAVKORGSLSYNPACGRFA----- 383
 QY 422 FDLVSVRNVSRNRQKADGLSSRHELLVEIDSFPGLDITKNTGYGLADEAFA-- 479
 Db 384 -----EHADTFEPASSAPAVNOTERAAP-----HDDPAP 416
 QY 480 -----LIQHTRLYLNVNVIKELMAYQALCRGFENAIQLSPAPLQELLVWALKDE 534
 Db 417 VPESTVAEETSAYETSAAVDSFNHETA-----SNEQOYIAEETKITE 461
 QY 535 LMSDEKDEKLEIAEYNTLLEKNAEMINEFSIHIDQDKLTRLPVLDQYTPMDRLP 594
 Db 462 VPDS--PSLRSEADEVFOEEMKLPPEFDNNSATSOBELQALNKLEERP-TERFP 517
 QY 595 EFVLTALGNDYTWDDKCEGFRVASAVGNFYALHPILPNPNSGNIHLYKKNRDSM--ADE 652
 Db 518 E-----LETFQOMH-----GTYLFAQSKDGLFTIDQ 543
 QY 653 HAENDLISDE-----NDVQDELL 670
 Db 544 HAAQERIKYEFREKIGEVSDDLQDELL 570

RESULT 11

US-09-815-242-12077
 : Sequence 12077, Application US/09815242
 : Patent No. US20020061569A1
 : GENERAL INFORMATION:
 : APPLICANT: Haselbeck, Robert
 : APPLICANT: Ohlsen, Karl L.
 : APPLICANT: Zyskind, Judith W.
 : APPLICANT: Wall, Daniel
 : APPLICANT: Trawick, John D.
 : APPLICANT: Carr, Grant J.
 : APPLICANT: Yamamoto, Robert T.
 : APPLICANT: Xu, H. Howard
 : TITLE OF INVENTION: Identification of Essential Genes in
 : FILE OF INVENTION: Prokaryotes
 : FILE REFERENCE: ELITRA.011A
 : CURRENT FILING DATE: 2001-03-21
 : PRIOR FILING DATE: 2000-03-21
 : PRIOR FILING DATE: 2000-03-21
 : PRIOR FILING DATE: 2000-05-23
 : PRIOR FILING DATE: 2000-05-23
 : PRIOR FILING DATE: 2000-05-26
 : PRIOR FILING DATE: 2000-05-26
 : PRIOR FILING DATE: 2000-10-23
 : PRIOR FILING DATE: 2000-10-23
 : PRIOR FILING DATE: 2000-11-27
 : PRIOR FILING DATE: 2000-11-27
 : PRIOR FILING DATE: 2000-12-22
 : PRIOR FILING DATE: 2000-12-22
 : PRIOR FILING DATE: 2001-02-16
 : NUMBER OF SEQ ID NOS: 14110
 : SOFTWARE: FASTSEQ for Windows Version 4.0
 : SEQ ID NO 12077
 : LENGTH: 633
 : TYPE: PRF
 : ORGANISM: Pseudomonas aeruginosa
 : US-09-815-242-12077

Query Match 12.7%; Score 471.5; DB 10; Length 633;
 Best Local Similarity 33.1%; Pred. No. 2, 4e-28;
 Matches 112; Conservative 66; Mismatches 149; Indels 11; Gaps 4;


```

APPLICANT: Sass, Philip
APPLICANT: Kinzler, Kenneth
APPLICANT: Vogelstein, Bert
TITLE OF INVENTION: A method for generating hypermutable
TITLE OF INVENTION: Plants
FILE REFERENCE: 01107.00069
CURRENT APPLICATION NUMBER: US/09/749,601A
CURRENT FILING DATE: 2000-12-28
PRIOR APPLICATION NUMBER: 60/183,333
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 14
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12
LENGTH: 779
TYPE: prt
ORGANISM: Arabidopsis thaliana
US-09-749-601A-12

```

Query Match	10.8%;	Score 402;	DB 10;	Length 779;
Best Local Similarity	24.5%;	Pred. No. 7.9e-23;		
Matches 171;	Conservative 113;	Mismatches 267;	Indels 148;	Gaps 277;

```

0Y      4 P$PBGCGACPEPPRIKRLSESVNRIAGVIORE$ANKELEIENSIDAGASVSVAWKD    63
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
0Y      7 PSP-----TTTSPILRPNRNVIHRIOSGVOLLIDSNAKEIVENSIDAGATSIENRD    62
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      64 GGLKLIOVSDDGHI-----NFEDLALICERHTTSKRISAVEDIQITXSMGF   109
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      63 YEDFYQYIDNGCIGSIPTEFNFCVOQLIRPFYIAL---KHHTSKIEFTDLINTTYGFE   119
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      110 KGEALSMITYGHVTYTITFGOLHGVR$Y-BDGVMEDEPRKCAAVKETOVMENILEYN    168
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      120 REBALSSICALGNLTVERINNEVATLLIFDHSGLLTAEKTRARI GTTYTVAKLEPSN    178
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      169 MYARKKITL-ONSNDYPRKYDIYSRFVAVHHINTFSC-RKHGANRADY--H$A$ISSRID    224
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      179 LPVRSEFPERNIRIKREYGKLVSLNAAVALIAKVAFYC$NTTGNP$RSVANTOGROSLKD    238
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      225 AIRSYGASVADLLEIVSYEDAAD$IFKMDGYISANAYVAKKIIMILFIENDRLVDCTA    284
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      239 NIITYVGISTFSL-----QP$-TORNLADQ---YFINGRPYDMRK    277
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      285 LKRAIFEVSATLPQASKPEIYMSIHLPSEHAVDNHPTKKEYSLNOERKIIETINALE    344
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      278 VSLTVAMELKDI-S$RKYPVTLIDEITYGAGCALDNTVPDRKKRFFESDET$VIT$SIREGIN    336
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      345 EKLMSNTRTRIOTQALNISGAOANPKDKDVSEASMSGSGTSOKPIPVSOVRT-----    398
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      337 EIYSS$NASAIYVNRFEEN-----SQOPKAGVSSFOQRKSNIISGLYLDV$STRIGEMAI    391
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      399 DPNRPGSRHHTYWHGOSSMILK-KFDLVSVRNVRR$RKNODAGDLSRHHELLVEIDSSF    457
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      392 EKERP$-LREVINDSSPM$EKFEE-----IACGT$KGEGSL$V-HD-VTHLDKTP    440
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      458 HPGILDIVNKCTYVGLADEAPALLOHNTRLIYVNVNINSKELMOALOALCRGNFNALOLS    517
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      441 SKGPOL-NVT-----EKTDAO$KDIS$SSFAO-----                     468
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      518 EPAPLOGLLMWALKDELIMSDEKDEKTELIAEVNTEILLKENAMINEYFSIHIDODGKLT    577
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      469 --STLNTFVIMGRKHENIS-----TJISEPVLNRQSSRYR$KSF$EV    511
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      578 RLPAVLDOQYPPDDORLPEFLYA-----LGNDVYHDDEECRTRYA$AVGNFYA    625
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      512 R-ALASCIVEGDOLDMDWTISKEDMTPERDESEIGNRISPEQTQADNERHERVLYGOF-    567
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      626 LHAPLIPSP$NGIHLK$KRRND-SMADEAHENDLSDEN    663
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      568 -----NL$BTI$KLERDLEIYDQHADEK$FEEH    596

```

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 15:45:41 ; Search time 40 Seconds

(without alignments)
2411.836 Million cell updates/sec

Title: US-09-954-950-2

Perfect score: 724
Sequence: 1 MDESPRGCGCAGEPPRRIR.....DGFVVASLEKLYIFERC 724

Scoring table: OLIGO
Gapop 60.0 , Gapept 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

A_Geneseq_101002:*

- 1: /SID2/gcgdata/geneseq/genesep-emb1/AA1980.DAT:*
- 2: /SID2/gcgdata/geneseq/genesep-emb1/AA1981.DAT:*
- 3: /SID2/gcgdata/geneseq/genesep-emb1/AA1982.DAT:*
- 4: /SID2/gcgdata/geneseq/genesep-emb1/AA1983.DAT:*
- 5: /SID2/gcgdata/geneseq/genesep-emb1/AA1984.DAT:*
- 6: /SID2/gcgdata/geneseq/genesep-emb1/AA1985.DAT:*
- 7: /SID2/gcgdata/geneseq/genesep-emb1/AA1986.DAT:*
- 8: /SID2/gcgdata/geneseq/genesep-emb1/AA1987.DAT:*
- 9: /SID2/gcgdata/geneseq/genesep-emb1/AA1988.DAT:*
- 10: /SID2/gcgdata/geneseq/genesep-emb1/AA1989.DAT:*
- 11: /SID2/gcgdata/geneseq/genesep-emb1/AA1990.DAT:*
- 12: /SID2/gcgdata/geneseq/genesep-emb1/AA1991.DAT:*
- 13: /SID2/gcgdata/geneseq/genesep-emb1/AA1992.DAT:*
- 14: /SID2/gcgdata/geneseq/genesep-emb1/AA1993.DAT:*
- 15: /SID2/gcgdata/geneseq/genesep-emb1/AA1994.DAT:*
- 16: /SID2/gcgdata/geneseq/genesep-emb1/AA1995.DAT:*
- 17: /SID2/gcgdata/geneseq/genesep-emb1/AA1996.DAT:*
- 18: /SID2/gcgdata/geneseq/genesep-emb1/AA1997.DAT:*
- 19: /SID2/gcgdata/geneseq/genesep-emb1/AA1998.DAT:*
- 20: /SID2/gcgdata/geneseq/genesep-emb1/AA1999.DAT:*
- 21: /SID2/gcgdata/geneseq/genesep-emb1/AA2000.DAT:*
- 22: /SID2/gcgdata/geneseq/genesep-emb1/AA2001.DAT:*
- 23: /SID2/gcgdata/geneseq/genesep-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	724	100.0	724	23	AAE22978
2	19	2.6	737	22	AAE08710
3	19	2.6	737	23	AAE22979
4	14	1.9	752	18	AAW09036
5	14	1.9	756	16	AAE75785
6	14	1.9	756	14	AAE76071
7	14	1.9	756	16	AAE79008
8	14	1.9	756	22	AAE85854
9	14	1.9	756	22	AAE63956
10	14	1.9	756	23	AAE24359

11	14	1.9	756	23	AA98778	Human Mutl homolog
12	14	1.9	775	22	AAE75536	Human colon cancer
13	13	1.8	133	22	AAE85855	Human PMS2-134 pro
14	13	1.8	133	22	AAE63957	Amino acid sequenc
15	13	1.8	442	20	AAE37635	Protein involved i
16	13	1.8	580	20	AAE35452	Chlamydia pneumoni
17	13	1.8	900	22	AAE85851	Human PMS2 protein
18	13	1.8	931	16	AAE79009	Human DNA repair p
19	13	1.8	932	22	AAE85852	Human PMS1 protein
20	13	1.8	932	22	AAE63953	Amino acid sequenc
21	13	1.8	932	22	AAE63954	Amino acid sequenc
22	13	1.8	932	23	AAE24357	Human mismatch rep
23	13	1.8	932	23	AAE24357	Human mismatch rep
24	9	1.2	54	19	AAE44347	Human post meiotic
25	9	1.2	76	23	ABP06237	Human ORFX protein
26	9	1.2	326	23	ABP40089	Staphylococcus epi
27	9	1.2	363	22	AAE48199	Mature castor eny
28	9	1.2	391	13	AAE20011	Soybean seed stea
29	9	1.2	391	16	AAE82024	Stearyl-ACP-desat
30	9	1.2	391	19	AAE37939	Amino acid sequenc
31	9	1.2	396	12	AAE14188	C. tinctarius delt
32	9	1.2	396	13	AAE22048	Carthamus tinctori
33	9	1.2	396	15	AAE54033	Sequence encoding
34	9	1.2	396	19	AAE44349	Carthamus tinctori
35	9	1.2	396	19	AAE44350	Ricinus communis d
36	9	1.2	629	22	AAE35368	Haemophilus influe
37	9	1.2	629	22	AAE35368	Pseudomonas aerugi
38	9	1.2	664	22	AAE65767	Drosophila melanog
39	9	1.2	669	22	AAE33766	Staphylococcus aur
40	9	1.2	669	22	AAE36741	Staphylococcus aur
41	9	1.2	769	22	AAE85849	Yeast MLH1 protein
42	9	1.2	769	22	AAE63951	Amino acid sequenc
43	9	1.2	859	22	AAE85850	Mouse PMS2 protein
44	9	1.2	859	22	AAE63952	Amino acid sequenc
45	9	1.2	859	23	AAE24355	Mouse mismatch rep

ALIGNMENTS

RESULT 1	AAE22978	
ID	AAE22978	standard; Protein; 724 AA.
XX		
AC	AAE22978:	
XX		
DT	21-AUG-2002	(first entry)
XX		
DE	Rice MLH1 protein.	
XX		
KW	Rice; MLH1: gene mutation; mismatch repair system; transposon tagging; tissue-specific promoter; herbicidal safener; male sterility; detection;	
KW	pollen formation; base pair mismatch.	
XX		
OS	Oryza sativa.	
XX		
FT	Key	Location/Qualifiers
FT	Region	103..119
FT		/note="mutl/PMS1 signature sequence"
XX		
PN	WO200224890-A2.	
XX		
PD	28-MAR-2002.	
XX		
PF	18-SEP-2001; 2001MC-US29088.	
XX		
PR	18-SEP-2000; 2000US-233124P.	
XX		
PA	(PION-) PIONEER HI-BRED INT INC.	
XX		
PI	Mahajan PB;	
XX		
DR	WPI; 2002-416283/44.	

DR N-PSDB; AAD36728.

Novel rice MLH1 ortholog nucleic acid molecule for increasing efficiency of targeted gene mutation or homologous recombination in a plant and for generating plants with reversible male sterility

Claim 7; Fig 2; 90pp; English.

The invention relates to isolated rice MLH1 orthologue nucleic acids. The nucleic acid is useful for increasing the efficiency of targeted gene mutation or homologous recombination in a plant, by transforming a plant with expression cassette comprising the nucleic acid linked to a chemical inducible promoter, transforming the plant with nucleic acid comprising a sequence having a desired mutation or a sequence to be homologously recombined, where the transformation occurs in the presence of chemical compound capable of inducing the promoter and the plant's cellular mismatch repair system is inhibited and selecting the transformed plants that contained the mutation or homologously recombined nucleotide sequence. The plant cellular mismatch repair system is inhibited through the use of transposon tagging of an MLH1 gene, sense- and antisense-suppression of an MLH1 gene, antibody binding to an MLH1 polypeptide or its variant, and targeted mutagenesis of specific amino acid residues encoded by an MLH1 gene. The nucleic acid is also useful for producing reversible male sterility in a plant, by transforming a plant with an expression cassette comprising a lexa DNA binding site embedded in a tissue-specific promoter that drives expression in the plant operably linked to the nucleic acid when expressed disrupts pollen formation or function through inhibition of the plant's cellular mismatch repair system, transforming the plant with a second expression cassette comprising a nucleotide sequence encoding a lexa repressor protein operably linked to a chemically-inducible promoter that drives expression in the plant, and exposing the plant to a compound capable of inducing the chemical-inducible promoter, to induce expression of lexa repressor protein. The tissue-specific promoter is an anther-specific promoter and the chemical-inducible promoter is a herbicidal safener. The polypeptide encoded by the nucleic acid is useful for detecting, locating, or removing a base pair mismatch (SNP). The present sequence is rice MLH1 protein.

Sequence 724 AA:

Query Match 100.0%; Score 724; DB 23; Length 724;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 724; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 MDEPSRGGCGAGGPRIRIRLESYVNRITAGVIOBPSSAVKELIENSLDAGASSVYA 60
 1 MKDPSRGGCGAGGPRIRIRLESYVNRITAGVIOBPSSAVKELIENSLDAGASSVYA 60
 61 VKDGGKLIQVSDGDGIRFEDIALICERHTTSKLSAYEDLOTIKSMGRGELASMTYV 120
 61 VKDGGKLIQVSDGDGIRFEDIALICERHTTSKLSAYEDLOTIKSMGRGELASMTYV 120
 61 VKDGGKLIQVSDGDGIRFEDIALICERHTTSKLSAYEDLOTIKSMGRGELASMTYV 120
 121 GHVTTVTITBGLHGRVSTRGVMEKPCCAVAGTGVYVNFYNNVARKKTIONN 180
 121 GHVTTVTITBGLHGRVSTRGVMEKPCCAVAGTGVYVNFYNNVARKKTIONN 180
 121 GHVTTVTITBGLHGRVSTRGVMEKPCCAVAGTGVYVNFYNNVARKKTIONN 180
 121 GHVTTVTITBGLHGRVSTRGVMEKPCCAVAGTGVYVNFYNNVARKKTIONN 180
 181 DDVPKIVDFSRFAVHINVTFCRKHGANRADVHASTSSRDARSYVAGSVVDLIE 240
 181 DDVPKIVDFSRFAVHINVTFCRKHGANRADVHASTSSRDARSYVAGSVVDLIE 240
 181 DDVPKIVDFSRFAVHINVTFCRKHGANRADVHASTSSRDARSYVAGSVVDLIE 240
 181 DDVPKIVDFSRFAVHINVTFCRKHGANRADVHASTSSRDARSYVAGSVVDLIE 240
 241 IKVSYEDPADSIFPKMDGYISNANYVAKITMILFINDRLVDTALRAIEYVSATLPQA 300
 241 IKVSYEDPADSIFPKMDGYISNANYVAKITMILFINDRLVDTALRAIEYVSATLPQA 300
 241 IKVSYEDPADSIFPKMDGYISNANYVAKITMILFINDRLVDTALRAIEYVSATLPQA 300
 301 SKPFITMSIHLSBEHVDVNIHPKKEVSLNORRIETIRNATIEKLMNSNTRIRIOTQA 360
 301 SKPFITMSIHLSBEHVDVNIHPKKEVSLNORRIETIRNATIEKLMNSNTRIRIOTQA 360
 301 SKPFITMSIHLSBEHVDVNIHPKKEVSLNORRIETIRNATIEKLMNSNTRIRIOTQA 360
 361 LNTSGIAQANPQDKVSEASMGSGTSOKIPVSOAVTDPDRNSGRHTYWHQOSSNLEK 420
 361 LNTSGIAQANPQDKVSEASMGSGTSOKIPVSOAVTDPDRNSGRHTYWHQOSSNLEK 420
 361 LNTSGIAQANPQDKVSEASMGSGTSOKIPVSOAVTDPDRNSGRHTYWHQOSSNLEK 420

QY 421 KFDLVSVNVRNRRNOKDAGDLSRHELLVEIDSSFPGLDIYVNCYVGLADAPAL 480
 DB 421 KFDLVSVNVRNRRNOKDAGDLSRHELLVEIDSSFPGLDIYVNCYVGLADAPAL 480
 QY 481 IOHNTRLVNVVNISELMYQOALCRFGNFNAIQSEPAELOELLVMAIKDELMSDEK 540
 DB 481 IOHNTRLVNVVNISELMYQOALCRFGNFNAIQSEPAELOELLVMAIKDELMSDEK 540
 QY 541 DDEKLEIAEVTETELKRNAMINEYFSIHIDOGKTLRLPVVDOTPDMDRLPEVYAL 600
 DB 541 DDEKLEIAEVTETELKRNAMINEYFSIHIDOGKTLRLPVVDOTPDMDRLPEVYAL 600
 QY 601 GNDVTWDEKCEFTVSAVGNFYALHPITLPNPSGNGIHLVKNRDSMADEHAENDLIS 660
 DB 601 GNDVTWDEKCEFTVSAVGNFYALHPITLPNPSGNGIHLVKNRDSMADEHAENDLIS 660
 QY 661 DENDVDOLLAEMAAVNAQREWTIOHVLFPSSMRFLKPPSMATDGTFFVYASLEKLYKI 720
 DB 661 DENDVDOLLAEMAAVNAQREWTIOHVLFPSSMRFLKPPSMATDGTFFVYASLEKLYKI 720
 QY 721 FERC 724
 DB 721 FERC 724

RESULT 2

ID AAE08710 standard; Protein: 737 AA.

AAE08710;

15-NOV-2001 (first entry)

XX Arabidopsis thaliana PMS2 protein homologue MLH.

XX Hypermutable plant; dominant negative allele; mismatch repair gene;

XX WMR; cell line generation; PMS2; ATRMLH.

XX Arabidopsis thaliana.

XX MO200161012-AL.

XX 23-AUG-2001.

XX 28-DEC-2000; 2000WO-US35397.

XX 18-FEB-2000; 2000US-0183333.

XX (NICO/) NICOLAIDES N C.

XX (GRAS/) GRASSO L.

XX (SASS/) SASS P M.

XX (KINZ/) KINZLER K.

XX (VOGE/) VOGELSTEIN B.

XX Nicolaides NC, Grasso L, Sass PM, Kinzler K, Vogelstein B;

XX WPI; 2001-529913/58.

XX Example 1; Page 57-59; 72pp; English.

XX The invention relates to a method for generating hypermutable cell.

XX The method involves introducing into a plant cell a polynucleotide.

XX Comprising a dominant negative allele of a mismatch repair (MMR) gene.

XX The method is useful for generating hypermutable plants, new cell lines

XX and plant varieties. This is particularly useful for agriculturally

XX important crops. The method is also useful for generating crop plants

XX with new output traits and plant cells exhibiting new biochemicals for

XX commercial use. The present sequence is Arabidopsis thaliana (At)

CC MLH protein. This sequence is a homologue of MMR protein, PMS2.
XX
SQ Sequence 737 AA;

Query Match 2.6%; Score 19; DB 22; Length 737;
Best Local Similarity 100.0%; Pred. No. 1.7e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 20 RLEESVNNRIAGEVQRP 38
DB 31 RLEESVNNRIAGEVQRP 49

RESULT 3

ID AAE22979 standard; Protein; 737 AA.

XX AAE22979;

DT 21-AUG-2002 (first entry)

XX Arabidopsis thaliana MLH1 protein.

XX MLH1: gene mutation; mismatch repair system; transposon tagging;
KM tissue-specific promoter; herbicidal safener; male sterility; detection;
KM pollen formation; base pair mismatch.

XX Arabidopsis thaliana.

XX WO200224890-A2.

XX 28-MAR-2002.

XX 18-SEP-2001; 2001WO-US29088.

XX 18-SEP-2000; 2000US-233124P.

XX (PION-) PIONEER HI-BRED INT INC.

XX Mahajan PB;

XX WPI: 2002-416283/44.

XX N-PSDB: AAD36729.

XX Novel rice MLH1 ortholog nucleic acid molecule for increasing
PT efficiency of targeted gene mutation or homologous recombination in a
PT plant and for generating plants with reversible male sterility

XX Disclosure: Page 89-90; 90pp; English.

XX The invention relates to isolated rice MLH1 orthologue nucleic acids. The
CC nucleic acid is useful for increasing the efficiency of targeted gene
CC mutation or homologous recombination in a plant, by transforming a plant
CC with expression cassette comprising the nucleic acid linked to a chemical
CC inducible promoter, transforming the plant with nucleic acid comprising
CC a sequence having a desired mutation or a sequence to be homologously
CC recombined, where the transformation occurs in the presence of chemical
CC compound capable of inducing the promoter and the plant's cellular
CC mismatch repair system is inhibited and selecting the transformed plants
CC that contained the mutation or homologously recombined nucleotide
CC sequence. The plant cellular mismatch repair system is inhibited through
CC the use of transposon tagging of an MLH1 gene, sense- and antisense-
CC suppression of an MLH1 gene, antibody binding to an MLH1 polypeptide or
CC its variant, and targeted mutagenesis of specific amino acid residues
CC encoded by an MLH1 gene. The nucleic acid is also useful for producing
CC reversible male sterility in a plant, by transforming a plant with an
CC expression cassette comprising a lexA DNA binding site embedded in a
CC tissue-specific promoter that drives expression in the plant operably
CC linked to the nucleic acid when expressed disrupts pollen formation or
CC function through inhibition of the plant's cellular mismatch repair
CC system, transforming the plant with a second expression cassette
CC comprising a nucleotide sequence encoding a lexA repressor protein
CC operably linked to a chemically-inducible promoter that drives expression

CC in the plant, and exposing the plant to a compound capable of inducing
CC the chemical-inducible promoter, to induce expression of lexA repressor
CC protein. The tissue-specific promoter is an anther-specific promoter
CC and the chemical-inducible promoter is a herbicidal safener. The
CC polypeptide encoded by the nucleic acid is useful for detecting,
CC localizing, or removing a base pair mismatch (SNP). The present sequence
CC is Arabidopsis thaliana MLH1 protein.
XX

SQ Sequence 737 AA;

Query Match 2.6%; Score 19; DB 23; Length 737;
Best Local Similarity 100.0%; Pred. No. 1.7e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 20 RLEESVNNRIAGEVQRP 38
DB 31 RLEESVNNRIAGEVQRP 49

RESULT 4

ID AAM09036 standard; protein; 752 AA.

XX AAM09036;

DT 01-OCT-1997 (first entry)

XX Mismatch repair protein, hPMS2.

XX Human; mismatch repair protein; hMSH2; hMLH1; hPMS2; antibody; Ab;
KM proliferation; anti-cancer therapy; neoplastic tissue; DNA replication;
KM DNA repair; mutation; sporadic colorectal carcinoma; cancer; HNPCC;
KM inherited; non-polyposis colorectal carcinoma.

XX Homo sapiens.

XX WO9641192-A1.

XX 19-DEC-1996.

XX 07-JUN-1996; 96WO-US08957.

XX 07-JUN-1995; 95US-0480351.

XX (UJYO) UNIV JOHNS HOPKINS.

XX Burrell M, Hill DE, Kinzler KW, Vogelstein B;

XX WPI: 1997-077287/07.

XX Discriminating proliferating from non-proliferating cells in tissue
PT - using antibodies specifically immuno-reactive with mismatch repair
PT protein, esp. human MSH2

XX Disclosure: Page 23-25; 37pp; English.

XX The sequences given in AAM09034-36 represent the human mismatch repair
CC proteins, hMSH2, hMLH1 and hPMS2. In the method of the invention, these
CC proteins were identified by reaction with an antibody (Ab) specific for
CC them, therefor discriminating proliferating from non-proliferating cells
CC The method may be used for monitoring the effectiveness of anti-cancer
CC therapy in neoplastic tissue, by comparing the amount of Ab-Ag complexes
CC in the sample with an amount determined at an earlier time, in which a
CC reduction in the amount indicates an effective therapy. The Ab are
CC especially specifically immunoreactive with the MSH2 mismatch repair
CC gene, which is 1 of at least 4 genes encoding proteins involved in the
CC repair of mismatched nucleotides following DNA replication or repair.
CC Mutations in the MSH2 gene contribute to the development of sporadic
CC colorectal carcinoma, while germline MSH2 mutations are responsible for
CC approx. 50% of inherited, non-polyposis colorectal carcinoma (HNPCC).
CC Since MSH2 is ubiquitously expressed, development of other cancers are
CC also susceptible to alterations in MSH2.

SO Sequence 752 AA;
 Query Match 1.9%; Score 14; DB 18; Length 752;
 Best Local Similarity 100.0%; Pred. No. 3.1e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 25 VVNR1AGEV10RP 38
 DB 15 VVNR1AGEV10RP 28
 RESULT 5
 ID AAR75785 standard; Protein: 756 AA.
 AC AAR75785;
 DT 04-MAR-1996 (first entry)
 DE Human wild type MHL1, a MHL1 homologue.
 DE MHL1; wild type; MHL1 homologue; cancer diagnosis; mismatch repair;
 KM tumour; susceptibility; mutation detection.
 OS Homo sapiens.
 PI W09516793-A1.
 PN 22-JUN-1995.
 PD 16-DEC-1994; 94MO-US14746.
 PF 09-DEC-1994; 94US-035902.
 PR 17-DEC-1993; 93US-0168877.
 PR 08-MAR-1994; 94US-0209521.
 XX (DAND) DANA FARBER CANCER INST INC.
 PA (UYOR-) UNIV OREGON HEALTH SCI.
 PI Baker SM, Bollag RJ, Bronner CE, Kolodner RD, Liskey RM;
 DR WPI: 1995-231583/30.
 DR N-PSDB; AAO90814.
 XX
 PT Determn. of a mutation in a MHL1 homologue or gene prod. in a tissue
 PT - used to diagnose cancer susceptibility, and to identify and
 PT classify a DNA mismatch-repair-defective tumour
 PS Claim 33; Fig 3; 168pp; English.
 CC AAO90814 encodes AAR75785 the wild type MHL1, a MHL1 homologue. A
 CC mutation in a MHL1 or hPMST nucleic acid isolated from a subject,
 CC can be detected by comparing it with an analogous segment of the
 CC above wild type allele. This method can be used to diagnose cancer
 CC susceptibility, or to identify and classify a DNA mismatch-repair
 CC defective tumour.
 SO Sequence 756 AA;
 Query Match 1.9%; Score 14; DB 16; Length 756;
 Best Local Similarity 100.0%; Pred. No. 3.1e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 25 VVNR1AGEV10RP 38
 DB 15 VVNR1AGEV10RP 28
 RESULT 6
 ID AAR76071 standard; Protein: 756 AA.
 AC AAR76071;
 DT 03-AUG-1995.

XX 25-JAN-1996 (first entry)
 DE Human mismatch repair pathway protein, MHL1.
 DE Mismatch repair; MSH2; primer; identification; defect; alteration;
 KM Cancer; tumour; vaccine.
 OS Homo sapiens.
 PI W09514085-A2.
 PN 26-MAY-1995.
 PD 27-NOV-1994; 94MO-US13385.
 PF 13-JUN-1994; 94US-0259310.
 PR 27-NOV-1993; 93US-0154792.
 PR 07-DEC-1993; 93US-0163449.
 XX (UYVE-) UNIV VERMONT & STATE AGRIC COLLEGE.
 PA (DAND) DANA FARBER CANCER INST.
 PI Fishel R, Kolodner RD, Reenan RAG;
 DR WPI: 1995-200377/26.
 DR N-PSDB; AAO94015.
 XX
 PT Determining alteration in human mismatch repair pathways - used in
 PT the diagnosis, prognosis and therapy of cancers and in screening
 PT assays
 PS Example 13; Page 239-242; 256pp; English.
 CC AAR76071 is the human mismatch repair pathway protein MHL1. Defects or
 CC alterations in a human mismatch repair gene results in the accumulation
 CC of unstable repeated DNA sequences, a feature of a number of different
 CC cancers. The identification of a defect in the mismatch repair pathway
 CC can be diagnostic of a predisposition to cancer and prognostic for a
 CC particular mammalian cancer e.g colorectal, ovarian, endometrial
 CC (uterine), renal, bladder, skin, rectal and bowel. The nucleotide
 CC sequences and polypeptides of the hMSH2 gene may also be used for
 CC therapy and in vaccines.
 SO Sequence 756 AA;
 Query Match 1.9%; Score 14; DB 16; Length 756;
 Best Local Similarity 100.0%; Pred. No. 3.1e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 25 VVNR1AGEV10RP 38
 DB 15 VVNR1AGEV10RP 28
 RESULT 7
 ID AAR79008 standard; Protein: 756 AA.
 AC AAR79008;
 DT 26-MAR-1996 (first entry)
 DE Human DNA repair protein MHL1.
 DE DNA repair protein; MHL1; MHL2; MHL3; therapy; cancer; vectors;
 KM DNA synthesis; diagnosis; disease; multA.
 OS Homo sapiens.
 PI W09520678-A1.
 PN 03-AUG-1995.


```
XX 25-JAN-1995; 95WO-US01035.
PF 23-AUG-1994; 94US-0294312.
PR 27-JAN-1994; 94US-0187757.
PR 16-MAR-1994; 94US-0210143.
XX (HUMA-) HUMAN GENOME SCI INC.
PA Adams MD, Fleischmann RD, Fraser CM, Fulmer RA;
PI Haseltine WA, Kirtness EF, Rosen CA, Ruben SM, Wei Y;
DR WPI: 1995-275461/36.
XX N-PSDB; AAO97525.
XX Polynucleotide(s) encoding human mutL4 homologues, hMLH1, hMLH2 and
PT hMLH3 - used for therapeutic treatment of, e.g. hereditary cancer
XX Claim 8; Figure 1; 124pp: English.
XX The polynucleotides described in AAO97525-27 encode the human
CC analogues of the prokaryotic mutL4 DNA repair gene. The polypeptides
CC they encode (AA879008-R79010) are used for therapeutic purposes e.g.
CC in the treatment of cancer, esp. hereditary cancer. They may also
CC be used for in vitro manipulation of DNA, synthesis of DNA and the
CC manufacture of DNA vectors and in methods of diagnosing a disease or
CC a susceptibility to a disease related to a mutation in the hMLH1, -2
CC or -3 DNA repair genes.
XX
SQ Sequence 756 AA;
Query Match 1.9%; Score 14; DB 16; Length 756;
Best Local Similarity 100.0%; Pred. No. 3.1e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 25 VVNR1AGEV1ORP 38
DB 15 VVNR1AGEV1ORP 28
RESULT 8
AAB85854
ID AAB85854 standard; Protein; 756 AA.
XX AAB85854;
AC AAB85854;
XX
DE 29-OCT-2001 (first entry)
XX
DE Human MLH1 protein.
XX
XX Hypermutable bacteria; mismatch repair gene; MMR gene; MutH; MutS; MutL;
XX MutY; PMS2; MLH1; MLH3; PMSR; biocatalysis; bioremediation; biochemical;
XX drug discovery; detoxification; toxin; biotransformation.
XX
OS Homo sapiens.
XX
XX WO200159092-A2.
XX
XX 16-AUG-2001.
XX
XX 12-FEB-2001; 2001WO-US04339.
XX
XX 11-FEB-2000; 2000US-0181929.
XX
XX (UYJO ) UNIV JOHNS HOPKINS.
XX
XX Nicolaides NC, Sass PM, Grasso L, Vogelstein B, Kinzler KW;
XX WPI: 2001-514664/56.
XX N-PSDB; AAH76367.
XX
PT Making hypermutable bacteria for biocatalysis, bioremediation and drug
discovery, involves introducing polynucleotide comprising dominant
```

```
PT negative allele of mismatch repair gene under regulatory sequence
PT control
XX
XX Example 1; Page 44; 68pp: English.
XX
XX The invention provides a method for generating a hypermutable bacteria.
CC The method involves introducing a polynucleotide having a dominant
CC negative allele of a mismatch repair (MMR) gene under the control of an
CC inducible transcription regulatory sequence, into a bacterium. The cell
CC becomes inducibly hypermutable. The method is useful to create desirable
CC output traits for commercial applications, using dominant negative
CC alleles of mismatch repair proteins. The mismatch repair gene is a MutH,
CC MutS, MutL or MutR homologue and can be selected from PMS2, MLH1, MLH3,
CC PMSR or PMSR homologue. The hypermutable bacteria is useful for the
CC production, biocatalysis, bioremediation and drug discovery. It is also
CC useful in manufacturing industry for the generation of new biochemicals
CC useful for detoxifying noxious chemicals from by-products of
CC manufacturing processes or those used as catalysts, for remediation of
CC toxins present in the environment including polychlorobenzene, heavy
CC metals and other environmental hazards for which there is a need to
CC remove them from the environment. The hypermutable bacteria is also
CC useful for screening novel mutations in a gene or a set of genes that
CC produce variant siblings that exhibit a new output trait not found in
CC wild type cells. The bacteria are also useful for producing increased
CC quantity or quality of protein or non-protein therapeutic molecule e.g.
CC penicillin G, Erythromycin and Clavulanic acid, by biotransformation.
CC Dominant negative alleles of the MMR gene are useful for producing higher
CC quantities of recombinant polypeptides. The present sequence represents
CC a human MLH1 protein.
XX
SQ Sequence 756 AA;
Query Match 1.9%; Score 14; DB 22; Length 756;
Best Local Similarity 100.0%; Pred. No. 3.1e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 25 VVNR1AGEV1ORP 38
DB 15 VVNR1AGEV1ORP 28
RESULT 9
AAG63956
ID AAG63956 standard; Protein; 756 AA.
XX AAG63956;
AC AAG63956;
XX
DE 29-OCT-2001 (first entry)
XX
DE Amino acid sequence of human mismatch repair protein MLH1.
XX
XX MLH1; mismatch repair gene; MMR gene; hypermutable yeast.
XX
XX Homo sapiens.
XX
XX WO200162945-A1.
XX
XX 30-AUG-2001.
XX
XX 21-FEB-2001; 2001WO-US05447.
XX
XX 23-FEB-2000; 2000US-0184336.
XX
XX (UYJO ) UNIV JOHNS HOPKINS.
XX
XX (NICO/) NICOLAIDES N C.
XX
XX (SASS/) SASS P M.
XX
XX (GRAS/) GRASSO L.
XX
XX (VOGE/) VOGELSTEIN B.
XX
XX (KINZ/) KINZLER K W.
XX
XX Nicolaides NC, Sass PM, Grasso L, Vogelstein B, Kinzler KW;
XX WPI: 2001-52820/57.
XX
```

DR N-PSDB: AAH75044.
XX Making hypermutable yeast that exhibit novel selected output traits for
PT commercial applications, comprises introducing polynucleotide
PT containing dominant negative allele of mismatch repair gene
PS Disclosure: Page 43-44; 60pp; English.
XX The present sequence represents human MLH1. MLH1 is a mismatch repair
CC (MMR) gene. The specification describes a method for making a
CC hypermutable yeast, comprising introducing a polynucleotide containing
CC a dominant negative allele of a mismatch repair (MMR) gene, into a
CC yeast, whereby the cell becomes hypermutable. The method is useful
CC to create desirable output traits for commercial applications, using
CC dominant negative alleles of mismatch repair proteins. The hypermutable
CC yeast is useful for production, biocatalysis, bioremediation and drug
CC discovery. It is also useful in genetic screens for the direct selection
CC of variant subclones that exhibit new output traits. The hypermutable
CC yeast is also useful in the manufacturing industry for the generation
CC of new biochemicals, for detoxifying noxious chemicals from by-products
CC of manufacturing processes or those used as catalysts, for remediation
CC of toxins present in the environment including polychlorobenzenes, heavy
CC metals and other environmental hazards for which there is a need to
CC remove them from the environment. The yeast is also useful for producing
CC increased quantity or quality of protein or non-protein therapeutic
CC molecule e.g., penicillin G, erythromycin and clavulanic acid, by
CC biotransformation.
XX Sequence 756 AA:
SO
Query Match 1.9%; Score 14; DB 22; Length 756;
Best Local Similarity 100.0%; Pred. No. 3.1e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 25 VVNR1AGEV1ORP 38
DB 15 VVNR1AGEV1ORP 28
RESULT 10
AAE24359
ID AAE24359 standard; Protein; 756 AA.
XX AAE24359:
AC 04-OCT-2002 (first entry)
DT Human mismatch repair protein, MLH1.
DE Human mismatch repair protein, MLH1.
KW Hypermutable: antibody-producing cell; dominant negative allele; MMR;
KW mismatch repair gene; genetic diversity; biochemical activity; human.
XX Homo sapiens.
OS
XX WO200237967-A1.
PN 16-MAY-2002.
PD
XX 07-NOV-2000; 2000WO-US30588.
PF
XX 07-NOV-2000; 2000WO-US30588.
PR (MORP-) MORPHOTEX INC.
PA Nicolaides NC, Grasso L, Sass PM;
XX WPI: 2002-479786/51.
DR N-PSDB: AAD39200.
XX Making hypermutable antibody-producing cells for producing antibodies
PT with e.g. enhanced biochemical activity, comprises introducing into a
PT cell a polynucleotide with a dominant negative allele of a mismatch
PT repair gene

XX Disclosure: Page 18; 75pp; English.
PS The invention relates to a method for making a hypermutable, antibody-
XX producing cell. The method comprises introducing into a cell, which is
CC capable of producing antibodies, a polynucleotide comprising a dominant
CC negative allele of a mismatch repair gene (MMR). The method is useful
CC for generating genetically altered antibody-producing cell lines with
CC improved antibody characteristics. In particular, the method is useful
CC for generating genetic diversity within immunoglobulin genes directed
CC against an antigen to produce antibodies with enhanced biochemical
CC activity or for generating antibody-producing cells with increased level
CC of antibody production. The present sequence is human mismatch repair
CC protein, MLH1.
XX Sequence 756 AA:
SO
Query Match 1.9%; Score 14; DB 23; Length 756;
Best Local Similarity 100.0%; Pred. No. 3.1e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 25 VVNR1AGEV1ORP 38
DB 15 VVNR1AGEV1ORP 28
RESULT 11
AAU98778
ID AAU98778 standard; Protein; 756 AA.
XX AAU98778:
AC 24-SEP-2002 (first entry)
DT Human Mutl homologue, MLH1.
DE Human Mutl homologue, MLH1.
KW Human, DNA mismatch repair enzyme; MLH1;
KW Mut L homologue; dominant negative mutation; antibacterial;
KW hypermutable cell; Post meiotic segregation increased; PMS1; PMS2;
KW MutS homologue 2; MSH2; MLH2; antimicrobial.
XX Homo sapiens.
OS
XX WO200238750-A1.
PN 16-MAY-2002.
PD
XX 07-NOV-2000; 2000WO-US30587.
PF
XX 07-NOV-2000; 2000WO-US30587.
PR (MORP-) MORPHOTEX INC.
PA Grasso L, Nicolaides NC, Sass PM;
XX WPI: 2002-508210/54.
DR N-PSDB: ABR86091.
XX Making mammalian cell hypermutable for obtaining a mammalian cell that
PT is resistant to selected microbe by introducing polynucleotide
PT comprising dominant-negative allele of mismatch repair gene into
PT mammalian cell
PS Disclosure: Page 18; 68pp; English.
XX The invention relates to making (M1) a mammalian cell hypermutable.
CC comprising introducing a polynucleotide with a dominant-negative allele
CC of a DNA mismatch repair gene (e.g. Post meiotic segregation increased 1
CC and 2 (PMS1 and PMS2), MutS homologue 2 (MSH2) and MutL homologue 1 and 2
CC (MLH1, 2) into the mammalian cell which is hypermutable. Also included is
CC a homogenous composition comprising a cultured, hypermutable, mammalian
CC cell with a dominant negative allele of mismatch repair gene.
CC (M1) is useful for making a mammalian cell hypermutable. The

CC composition is useful for obtaining a mammalian cell that is resistant
 CC to a selected microbe. (M1) Involves growing a culture of mammalian
 CC cells with a dominant negative allele of a mismatch repair gene,
 CC exposing the cells to the selected microbe, and selecting the mammalian
 CC cell resistant to the selected microbe. The hypermutable cell is
 CC resistant to gram-negative and gram-positive microbe, protozoan,
 CC bacteria or fungi. The microbial resistance is selected by isolating
 CC and testing conditioned medium from the hypermutable cell. The
 CC composition is also useful for obtaining a cell comprising a mutation
 CC in a gene encoding an antimicrobial activity. (M1) Involves growing a
 CC culture of mammalian cells with the gene encoding the antimicrobial
 CC activity, and a dominant negative allele of a mismatch repair gene,
 CC selecting a cell comprising the antimicrobial activity, and determining
 CC whether the gene comprises a mutation. The cell is examined to
 CC determine whether the gene comprises a mutation by analysing a
 CC nucleotide sequence of the gene or mRNA transcribed from the gene, a
 CC protein encoded by gene or its phenotype. The composition
 CC is useful for identifying antimicrobial agents, microbe-specific toxic
 CC molecules, and for producing new phenotypes of the cell. (M1) Is useful
 CC for creating genetically altered antimicrobial molecules, and also for
 CC creating cell lines that manufacture antimicrobial molecules for use in
 CC large scale production of antimicrobial agents for clinical studies.
 CC (M1) is also useful in cell lines that express known antimicrobial
 CC agents to enhance the biochemical activity of the antimicrobial agent.
 CC The present sequence represents human MLH1.

XX SQ Sequence 756 AA;

Query Match 1.9%; Score 14; DB 23; Length 756;
 Best Local Similarity 100.0%; Pred. No. 3.1e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 25 VVNR1AGEV1QRP 38
 Db 15 VVNR1AGEV1QRP 28

RESULT 12
 AAG75536
 ID AAG75536 standard; Protein; 775 AA.

XX AC AAG75536;

XX DT 03-SEP-2001 (first entry)

XX DE Human colon cancer antigen protein SEQ ID NO:6300.

XX KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
 XX colorectal carcinoma; chromosome 3.

XX OS Homo sapiens.

XX PN WO200122920-A2.

XX PD 05-APR-2001.

XX PE 28-SEP-2000; 2000WO-US26524.

XX PR 29-SEP-1999; 99US-0157137.

XX PR 03-NOV-1999; 99US-0163280.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;

XX DR WPI: 2001-235357/24.

XX DR N-PSDB; AAH34941.

XX PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 XX useful for preventing, diagnosing and/or treating colorectal cancers -
 XX Claim 11; Page 7756-7760; 9803pp; English.

CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or to supplement the patient's own production of P.
 CC Additionally, N may be used to produce the colon cancer-associated P,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAB77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX SQ Sequence 775 AA;

Query Match 1.9%; Score 14; DB 22; Length 775;
 Best Local Similarity 100.0%; Pred. No. 3.2e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 25 VVNR1AGEV1QRP 38
 Db 34 VVNR1AGEV1QRP 47

RESULT 13
 AAB85855
 ID AAB85855 standard; Protein; 133 AA.

XX AC AAB85855;

XX DT 29-OCT-2001 (first entry)

XX DE Human PMS2-134 protein.

XX KW Hypermutable bacteria; mismatch repair gene; MMR gene; Muth; Muts; Mutt;
 XX Muty; PMS2; MLH1; MLH3; PMSR; biocatalysis; bioremediation; biochemical;
 XX drug discovery; detoxification; toxin; biotransformation; PMS2-134.

XX OS Homo sapiens.

XX PN WO200159092-A2.

XX PD 16-AUG-2001.

XX PE 12-FEB-2001; 2001WO-US04339.

XX PR 11-FEB-2000; 2000US-0181929.

XX PA (UYJO) UNIV JOHNS HOPKINS.

XX PI Nicolaides NC, Sass PM, Grasso L, Vogelstein B, Kinzler KW;

XX DR WPI: 2001-514664/56.

XX DR N-PSDB; AAH76368.

XX PT Making hypermutable bacteria for biocatalysis, bioremediation and drug
 XX discovery, involves introducing polynucleotide comprising dominant
 XX negative allele of mismatch repair gene under regulatory sequence
 XX control

XX PT Example 1; Page 45; 68pp; English.

XX PS The invention provides a method for generating a hypermutable bacteria.

XX CC The method involves introducing a polynucleotide having a dominant
 XX negative allele of a mismatch repair (MMR) gene under the control of an
 XX inducible transcription regulatory sequence, into a bacterium. The cell

becomes inducibly hypermutable. The method is useful to create desirable output traits for commercial applications, using dominant negative alleles of mismatch repair proteins. The mismatch repair gene is a MTH, Muts, Muls or Multy homologue and can be selected from PMS2, MHL, MHS, PMSR or PMSR homologue. The hypermutable bacteria is useful for the production, biocatalysis, bioremediation and drug discovery. It is also useful in manufacturing industry for the generation of new biochemicals useful for detoxifying noxious chemicals from by-products of manufacturing processes or those used as catalysts, for remediation of toxins present in the environment including polychlorobenzenes, heavy metals and other environmental hazards for which there is a need to remove them from the environment. The hypermutable bacteria is also useful for screening novel mutations in a gene or a set of genes that produce variant siblings that exhibit a new output trait not found in wild type cells. The bacteria are also useful for producing increased quantity or quality of protein or non-protein therapeutic molecule e.g. Penicillin G, Erythromycin and Clavulanic acid, by biotransformation. Dominant negative alleles of the MMR gene are useful for producing higher quantities of recombinant polypeptides. The present sequence represents a human PMS2-134 protein.

Sequence 133 AA:

Query Match 1.8%; Score 13; DB 22; Length 133;

Best Local Similarity 100.0%; Pred. No. 6.4e-05; Mismatches 0; Indels 0; Gaps 0;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 42 VKELIENSLDAGA 54

DB 25 VKELIENSLDAGA 37

RESULT 14

ID AAG63957 standard; Protein; 133 AA.

AC AAG63957;

DT 29-OCT-2001 (first entry)

DE Amino acid sequence of human mismatch repair protein PMS2-134.

KW PMS2-134; mismatch repair gene; MMR gene; hypermutable yeast.

OS Homo sapiens.

PN WO200162945-A1.

PD 30-AUG-2001.

PF 21-FEB-2001; 2001WO-US05447.

PR 23-FEB-2000; 2000US-0184336.

PA (UYJO) UNIV JOHNS HOPKINS.

PA (NICO/) NICOLAIDES N C.

PA (SASS/) SASS P M.

PA (GRAS/) GRASSO L.

PA (VOGE/) VOGELSTEIN B.

PA (KINZ/) KINZLER K W.

PI Nicolaides NC, Sass PM, Grasso L, Vogelstein B, Kinzler KW.

DR WPI; 2001-522820/57.

DR N-PSDB; AAH75045.

XX Making hypermutable yeast that exhibit novel selected output traits for commercial applications, comprises introducing polynucleotide containing dominant negative allele of mismatch repair gene

PS Disclosure; Page 45; 60pp; English.

XX The present sequence represents human PMS2-134. PMS2-134 is a mismatch

repair (MMR) gene. The specification describes a method for making a hypermutable yeast, comprising introducing a polynucleotide containing a dominant negative allele of a mismatch repair (MMR) gene, into a yeast, whereby the cell becomes hypermutable. The method is useful to create desirable output traits for commercial applications, using dominant negative alleles of mismatch repair proteins. The hypermutable yeast is useful for production, biocatalysis, bioremediation and drug discovery. It is also useful in genetic screens for the direct selection of variant subclones that exhibit new output traits. The hypermutable yeast is also useful in the manufacturing industry for the generation of new biochemicals, for detoxifying noxious chemicals from by-products of manufacturing processes or those used as catalysts, for remediation of toxins present in the environment including polychlorobenzenes, heavy metals and other environmental hazards for which there is a need to remove them from the environment. The yeast is also useful for producing increased quantity or quality of protein or non-protein therapeutic molecule e.g., Penicillin G, Erythromycin and Clavulanic acid, by biotransformation.

Sequence 133 AA:

Query Match 1.8%; Score 13; DB 22; Length 133;

Best Local Similarity 100.0%; Pred. No. 6.4e-05; Mismatches 0; Indels 0; Gaps 0;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 42 VKELIENSLDAGA 54

DB 25 VKELIENSLDAGA 37

RESULT 15

ID AAY37635 standard; Protein; 442 AA.

AC AAY37635;

DT 07-OCT-1999 (first entry)

DE Protein involved in intermediate metabolism of nucleic acids.

KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma; Paratrachoma; inclusion conjunctivitis; genital disease; perlepatitis;

KW Pongonococcal urethritis; epididymitis; cervicitis; salpingitis; Bactrolinitis; pneumopathy; venereal lymphogranulomatosis.

OS Chlamydia trachomatis.

PN WO928475-A2.

PD 10-JUN-1999.

PF 27-NOV-1998; 98WO-1B01939.

PR 04-NOV-1998; 98US-0107077.

PR 28-NOV-1997; 97FR-0015041.

PR 17-DEC-1997; 97FR-0016034.

PA (GEST) GENSET.

PA Griffiths R;

PI WPI; 1999-371125/31.

DR Genome sequence of Chlamydia trachomatis

PT Disclosure; Page 1273-1274; 1755pp; English.

XX AAY37635-1737949 are encoded by open reading frames (ORFs) of the genome of Chlamydia trachomatis (see AA201425). The polypeptides can be used as vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences can also be used to control growth of the microorganism. Chlamydia trachomatis is responsible for a large number of diseases, e.g. eye diseases such as conventional trachoma, nonendemic trachoma,

CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
 CC perihepatitis, bartolinitis; pneumonia in breast feeding infants;
 CC and venereal lymphogranulomatosis. The polypeptides of the invention
 CC may be of use in treating these diseases.

XX
 SQ Sequence 442 AA;

Query Match 1.8%; Score 13; DB 20; Length 442;
 Best Local Similarity 100.0%; Pred. No. 0.00021;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 42 VKELIENSIDAGA 54
 |||||H|||||||
 DB 32 VKELIENSIDAGA 44

Search completed: March 27, 2003, 15:51:23
 Job Time : 42 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 15:42:51 ; Search time 26 Seconds
(without alignments)
2676.974 Million cell updates/sec

Title: US-09-954-950-2

Perfect score: 3709

Sequence: 1 MDEPSPRGCGAGGEPPIRR.....DGTFTVQVASELEKTKIERRC 724

Scoring table: BLOSUM62
Gapop 10.0 ; Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2505	67.5	737	2	MLH1 protein [limp
2	2505	67.5	737	2	DNA mismatch repair
3	1452	39.1	756	2	DNA mismatch repair
4	1097	29.6	769	2	mismatch repair pr
5	1040.5	28.1	684	2	probable DNA misa
6	845	22.8	779	2	hypothetical prote
7	597	16.1	669	2	DNA mismatch repair
8	584.5	15.8	622	2	DNA mismatch repair
9	565	15.2	649	2	DNA mismatch repair
10	565	15.2	649	2	DNA mismatch repair
11	560	15.1	649	2	DNA mismatch repair
12	555	15.0	695	2	DNA mismatch repair
13	547.5	14.8	629	2	mismatch repair pr
14	538	14.5	658	2	DNA mismatch repair
15	532	14.3	603	2	DNA mismatch repair
16	529	14.3	658	2	DNA mismatch repair
17	523	14.1	601	2	DNA mismatch repair
18	509.5	13.7	627	2	DNA mismatch repair
19	505	13.6	637	2	DNA mismatch repair
20	486.5	13.1	618	2	DNA mismatch repair
21	482.5	13.0	618	2	DNA mismatch repair
22	471.5	12.7	633	2	DNA mismatch repair
23	465.5	12.6	610	2	DNA mismatch repair
24	461	12.4	609	2	DNA mismatch repair
25	461	12.4	615	2	DNA mismatch repair
26	461	12.4	615	2	DNA mismatch repair
27	461	12.4	628	2	DNA mismatch repair
28	461	12.4	653	2	DNA mismatch repair
29	459	12.4	613	2	methyl-directed ml

30	459	12.4	635	2	AC0046	DNA mismatch repair
31	456.5	12.3	610	2	A97870	DNA mismatch repair
32	456	12.3	659	2	E84176	DNA mismatch repair
33	448	12.1	623	2	A13637	DNA mismatch repair
34	447.5	12.1	862	2	S47598	mult. protein homol
35	446.5	12.0	794	2	T37989	DNA mismatch repair
36	437.5	11.8	516	2	H72427	DNA mismatch repair
37	433.5	11.7	584	2	D84996	DNA mismatch repair
38	433.5	11.7	595	2	F71650	DNA mismatch repair
39	426.5	11.5	619	2	E82765	DNA mismatch repair
40	425.5	11.5	580	2	A73032	DNA mismatch repair
41	425.5	11.5	580	2	B86592	DNA mismatch repair
42	415.5	11.2	576	2	A71497	probable DNA misa
43	415.5	11.2	637	2	D87335	DNA mismatch repair
44	414	11.2	648	2	A71342	probable DNA misa
45	409.5	11.0	904	2	S53896	DNA mismatch repair

ALIGNMENTS

RESULT 1

F85092

MLH1 protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Mar-2001

C:Accession: F85092

R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp

Nature 402, 769-777, 1999

A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: A85001; MUID:20083488; PMID:10617198

A:Accession: F85092

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-737 <STO>

A:Cross-References: GB:NC_001268; NID:g7267557; PIDN:CAB78038.1; GSPDB:GN00140

A:Gene: AT4G09140

A:Map position: 4

C:Superfamily: mismatch repair protein MLH1

Query Match	Score	DB 2;	Length	737;
Best local Similarity	66.4%	Pred. No. 5.5e-146;		
Matches	482;	Conservative 111;	Mismatches 129;	Indels 4;
Gaps	3;			
OY	2	DEPSPRGCGAGGEPPIRRLEESVNRIRAGETIOPSSAVKELIENSIDAGASSVAV 61		
DB	13	EEESPATTTVPREPPIQRLSESVNRIRAGETIOPSSAVKELIENSIDAGASSVAV 72		
OY	62	KDGLKLIQVSDGHDGIRFEDLILCERHTTSKLISAYEDLOTIKSMGFGEALASMTYVG 121		
DB	73	KDGLKLIQVSDGHDGIRFEDLILCERHTTSKLISAYEDLOTIKSMGFGEALASMTYVG 132		
OY	122	HVTVTTFEGQLHGYNVSRDGVMEHPKCAAVKGTQVWENLFYNNVARKKTLQNSND 181		
DB	133	HVTVTTFEGQLHGYNVSRDGVMEHPKCAAVKGTQVWENLFYNNVARKKTLQNSND 192		
OY	182	DYKTIYDFISRAVHHINTFSCRKIGANRADYHASTSRDLAINSVGASVVRDIIET 241		
DB	193	DYKTIYDFISRAVHHINTFSCRKIGANRADYHASTSRDLAINSVGASVVRDIIET 252		
OY	242	KVSEYDAAGSIFFMDGYISNANVAKKTIPIILFNRLVYCTALKRAIEVYSATLPOAS 301		
DB	253	KVSEYDAAGSIFFMDGYISNANVAKKTIPIILFNRLVYCTALKRAIEVYSATLPOAS 312		
OY	302	KPEIYVSHLSEHVDVNIHPTKKEVSLNQERIETIRNAIEKILMSNTTRIPOTQAL 361		
DB	313	KPEIYVSHLSEHVDVNIHPTKKEVSLNQERIETIRNAIEKILMSNTTRIPOTQAL 372		
OY	362	NLSGIAQANPKQKVEASGSKTSOKIPVSOAVTDPNPGRGLHTYHHGSSNLEK 421		
DB	373	EYIQ-STLISQKSDSPVSKPQSKOKTQKVPVNMKVRTDSSDPAGRLHAFIQPKPQSLPK 431		

QY 422 FDLVS -VRNVRBRNOKADGLSSRHELVIEDSSPHGLLDIYKNTCTYGLADEAFAL 480
 DB 432 VSSLSVSSVRBRNOKADGLSSRHELVIEDSSPHGLLDIYKNTCTYGLADEAFAL 491
 QY 481 IOHNTRLIYLVNVIKELMVOALCFGNFNAIOLSEAPLOELLVMAKDEL -MSD 538
 DB 492 VOYNTHLIYLVNVIKELMVOALCFGNFNAIOLSEAPLOELLVMAKDEL -MSD 551
 QY 539 EKDEKLEIAEVNTEILKENAEMINEFESIHDODGKILRLPVLDQYTPMDRLPEFVL 598
 DB 552 TKDLEKLEIAEVNTEILKENAEMINEFESIHDODGKILRLPVLDQYTPMDRLPEFVL 611
 QY 599 ALGNDVTDWDEKECFRIVASAVNGFYALHPILPNSGNGIHLKKNRDSMADEHENDL 658
 DB 612 CIGNDVEMWEDKSCFOGVSAAIGNFYAMHPILPNSGNGIHLKKNRDSMADEHENDL 671
 QY 659 ISDENDVDOELLAEAAAOAGRENTIOHVLPFSPKRLFLKPKKSMATDGTFOVASLEKLY 718
 DB 672 VDMEDNDLDDLLSDAENMAOAGRENTIOHVLPFSPKRLFLKPKKSMATDGTFOVASLEKLY 731
 QY 719 KIFERC 724
 DB 732 KIFERC 737

RESULT 2

T51620
 DNA mismatch repair protein MLH1 (imported) - Arabidopsis thaliana (fragment)
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 18-Aug-2000 #sequence-revision 18-Aug-2000 #text-change 03-Nov-2000
 C:Accession: T51620
 R:Jenn, M.; Pelletier, J.; Hilpert, M.; Belzille, F.; Kunze, R.
 Mol. Gen. Genet. 262, 633-642, 1999
 A:Title: Isolation and characterization of AtMLH1, a MutL homologue from Arabidopsis the
 A:Accession: T51620
 A:Reference number: 225418
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-737 <JEAN>
 A:Cross-references: EMBL:AJ012747; PIDN:CAA10163.1
 C:Genetics:
 A:Gene: MLH1
 A:Map position: 4
 C:Superfamily: mismatch repair protein MLH1
 Query Match 67.5%; Score 2505; DB 2; Length 737;
 Best Local Similarity 66.4%; Pred. No. 5.5e-146; Indels 4; Gaps 3;
 Matches: 482; Conservative 111; Mismatches 129;

QY 2 DEPSPPGGCAGRPRIIRLESVVRNIAAGEVIOFPSAVKELIENSIDAGASSVAV 61
 DB 13 EEPSPATTIYPRRPKIOIRLESVVRNIAAGEVIOFPSAVKELIENSIDAGASSVAV 72
 QY 62 KDGGIKLIQVSDGHRIFEDLALICERTTSKLSAYEDLOTIKSGFGEALASMTYVG 121
 DB 73 KDGGIKLIQVSDGHRIFEDLALICERTTSKLSAYEDLOTIKSGFGEALASMTYVG 132
 QY 122 HVTYTTTTEGOLHGYRYVSYDGVMEHPKCAAVKGTQVAVENLFFNMAARKKTIONSD 181
 DB 133 HVTYTTTTEGOLHGYRYVSYDGVMEHPKCAAVKGTQVAVENLFFNMAARKKTIONSD 192
 QY 182 DYPKIVDFISRPVAVHINVTFSCKKHANADVASASTSRDLAIVSYGASVYDLIEI 241
 DB 193 DYPKIVDFISRPVAVHINVTFSCKKHANADVASASTSRDLAIVSYGASVYDLIEI 252
 QY 242 KYSYDADSDIFPKMDGYISNNANYAKKITMLFTINDRLVDTALAKRIEYVATLPOAS 301
 DB 253 KYSYDADSDIFPKMDGYISNNANYAKKITMLFTINDRLVDTALAKRIEYVATLPOAS 312
 QY 302 KPFIVYSHLPSEHVDVNIHPKRVASLNLNRIETIRNAIEKILNMSVTRITFOAL 361
 DB 313 KPFIVYSHLPSEHVDVNIHPKRVASLNLNRIETIRNAIEKILNMSVTRITFOAL 372

QY 362 NLSTGIANPORKDYSEASMGSTKQKIPYSONVWRTPDRNPSGRILTYHMGSSNIEK 421
 DB 373 EYIQ -STLSQSDSVSVSGKPSGOKTOKVYVKNVWRTPDRNPSGRILTYHMGSSNIEK 431
 QY 422 FDLVS -VRNVRBRNOKADGLSSRHELVIEDSSPHGLLDIYKNTCTYGLADEAFAL 480
 DB 432 VSSLSVSSVRBRNOKADGLSSRHELVIEDSSPHGLLDIYKNTCTYGLADEAFAL 491
 QY 481 IOHNTRLIYLVNVIKELMVOALCFGNFNAIOLSEAPLOELLVMAKDEL -MSD 538
 DB 492 VOYNTHLIYLVNVIKELMVOALCFGNFNAIOLSEAPLOELLVMAKDEL -MSD 551
 QY 539 EKDEKLEIAEVNTEILKENAEMINEFESIHDODGKILRLPVLDQYTPMDRLPEFVL 598
 DB 552 TKDLEKLEIAEVNTEILKENAEMINEFESIHDODGKILRLPVLDQYTPMDRLPEFVL 611
 QY 599 ALGNDVTDWDEKECFRIVASAVNGFYALHPILPNSGNGIHLKKNRDSMADEHENDL 658
 DB 612 CIGNDVEMWEDKSCFOGVSAAIGNFYAMHPILPNSGNGIHLKKNRDSMADEHENDL 671
 QY 659 ISDENDVDOELLAEAAAOAGRENTIOHVLPFSPKRLFLKPKKSMATDGTFOVASLEKLY 718
 DB 672 VDMEDNDLDDLLSDAENMAOAGRENTIOHVLPFSPKRLFLKPKKSMATDGTFOVASLEKLY 731
 QY 719 KIFERC 724
 DB 732 KIFERC 737

RESULT 3

S43085
 DNA mismatch repair protein MLH1 - human
 C:Species: Homo sapiens (man)
 C:Date: 20-Oct-1994 #sequence-revision 10-Nov-1995 #text-change 29-Sep-1999
 C:Accession: S43085; 137414; 137549
 R:Bronner, C.E.; Baker, S.M.; Morrison, P.T.; Warren, G.; Smith, L.G.; Lescoe, M.K.;
 Kido, M.; Fishel, R.; Kolodner, R.; Liskay, R.M.
 Nature 368, 258-261, 1994
 A:Title: Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with
 A:Reference number: S43085; MUID:94195398; PMID:8145827
 A:Molecule type: mRNA
 A:Accession: S43085
 A:Residues: 1-756 <RES>
 A:Cross-references: GB:007343; NID:9463988; PIDN:AA82079.1; PID:9463987
 R:Kolodner, R.D.; Hall, N.R.; Lipford, J.R.; Kane, M.F.; Morrison, P.; Finan, P.J.; B
 Cancer Res. 55, 242-248, 1995
 A:Title: Structure of the human MLH1 locus and analysis of a large hereditary nonpoly
 A:Reference number: 137414; MUID:95112274; PMID:7812952
 A:Accession: 137414
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-756 <RES>
 A:Cross-references: EMBL:U04978; NID:91079785; PIDN:AA82079.1; PID:91079787
 R:Papadopoulos, N.; Nicolaides, N.C.; Wei, Y.F.; Ruben, S.M.; Carter, K.C.; Rosen, C
 Science 263, 1625-1629, 1994
 A:Title: Mutation of a mutL homolog is associated with hereditary Colon Cancer.
 A:Reference number: AA9945; MUID:94174309; PMID:8128251
 A:Accession: 137549
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-756 <RES>
 A:Cross-references: EMBL:U07418; NID:9466461; PIDN:AA17374.1; PID:9466462
 C:Genetics:
 A:Gene: MLH1; COCA2; HNPCC
 A:Cross-references: GDB:249617; OMIM:120436
 A:Map position: 3p21.3-3p21.3
 A:Insertions: 39/2; 69/3; 102/3; 127/2; 151/3; 182/2; 196/3; 226/2; 264/1; 295/2; 346/3.
 C:Superfamily: mismatch repair protein MLH1
 C:Keywords: DNA repair

Query Match 39.1%; Score 1452; DB 2; Length 756;
 39.1%; Score 1452; DB 2; Length 756;

Best Local Similarity 39.6%; Pred. No. 2.7e-81;
Matches 311; Conservative 145; Mismatches 214; Indels 116; Gaps 11;

```

OY 18 IRRLEESVYNNRIAGEVYQIORSASAVKELIENSIDAGASSVAVKDGKLIQVSDGCH 77
Db 8 IRRDETVYNNRIAGEVYQIORSANAIKEMENCLDAKSTISQIVYVEGGKLIQIDNGNG 67
OY 78 IRRFEDALICEHNTSKISAYEDLOTISKMGREGALASMTYVGHVYVTTTEGOLHGR 137
Db 68 IRRKEDLVICERFTTSKIOSPELDAISISTYGRGALASISVAVHTTTKADGCAVR 127
OY 138 VSYRGVNEKERPCAAVKGTOVMEVNFYNNVARKKTLONSNDYPRKYVDSISRAVHH 197
Db 128 ASYSGKILAKPKPCAGNGQIOTIYEDLFYNTATRKALKNSEEGKILEVVGRIYVHN 187
OY 198 INVETSCRHGANRADVSHASTSRDLAIRSYGASVVDLIEVYSYEDADSIFFKMG 257
Db 188 AGISFSYKOGQETADVTLPRASVVDNIRISIFGNANSELIEIGCEKTLA---FKMG 244
OY 238 YISNANYAKKTTMLFINDRLDCTALKRAIEFVYSATLPQASKPFYMSIHLSEHVD 317
Db 245 YISNANYAKKICIFLFIHRLVESTSLRKAIEFYAAVLPKNTHPFLYLSLEISPOVD 304
OY 318 VHIETKKEVSLNORIEETIRNAIEEKLMSNTTRIFOTQALNSGIAQANPO---KDK 375
Db 305 VAVHPTKHEVHLHEESTILEROVQIHESKLGSSRMFTYTL-LPGLAGSGEYVNST 363
OY 376 VSEASMGSGTSQKIPVSOVMTDRPNPSGRILHTWHGSSNM-----EKKFELVS 426
Db 364 TSLTSSSTSGSDKYAHQMVTRDSREO---KIDATLQPLSKFLSSPOAIVYEDKTDISS 421
OY 427 VANNVR-----SRNOKDAGDS-----444
Db 422 GRAROODDEMELEPAPAEVAAKNOSLEGDTTGTSEMSSEKRGPTSPNKRHRREDSDVM 481
OY 445 -----SRHELLV-----EIDSPHPELDIVKNCYVGLADEAFALI 481
Db 482 VEDDSRKREKTACTPRRRINTLSTVLSLOEELNEOGHEVLRKMLNHSVGCYNPOMALA 541
OY 482 QHNTRLYLIVNANNISKELMYQALCRFGNFNAIOLSEPAIOLLELMALKDELDMSDKO 541
Db 542 QHQTLYLINTKLSBELFYOILYIDFANFVGLRSEPARPLDLMALADSESGTEED 601
OY 542 DEKLEIAEVNTEILKENAMINEYSIHIDOGKTLRLPVIDOTPDMDRLPEFYALG 601
Db 602 GKPEGLAEIVTEFLKKKEMELADYFSLIEDENGLIGLPLIDNIVYPPLEGIPIFLRLA 661
OY 602 NDVTMDDEKCEPRTVASAVNGFYALHPILPNSGNGHLYKKNDSMADEHAENDLSD 661
Db 662 TEVNMDEKECESLSEKCAFMTST-----RKQYISE 693
OY 662 ENDV---DOELIAEAEAQAQREWTIOHVLFPMSRLFLKPKSMATDGTFOVVASLEKLY 718
Db 694 ESTLSGQOSEVYGSIPNSM---KWIYEHLYKALRSHILPRHFTEDGNILOLANLPDLX 750
OY 719 KIFERC 724
Db 751 KYFERC 756

```

RESULT 4

S54525

mismatch repair protein MLH1 - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein YM8520.16; protein YMR167W

C:Species: Saccharomyces cerevisiae

C:Date: 08-Jul-1995 #sequence revision 01-Sep-1995 #text change 29-Sep-1999

C:Accession: S54525; S48890; S54612

R:Hunt, S.; Bowman, S.

submitted to the EMBL Data Library, May 1995

A:Reference number: S54510

A:Accession: S54525

A:Molecule type: DNA

A:Residues: 1-769 <HUN>

A:Cross-references: GB:Z49705; EMBL:Z49700; NID:9825556; PIDD:CA89803.1; PID:9825572
A:Experimental source: strain AB972
R:Prolla, T.A.; Christie, D.M.; Liskay, R.M.
Mol. Cell. Biol. 14, 407-415, 1994
A:Title: Dual requirement in yeast DNA mismatch repair for MLH1 and PMS1, two homolog
A:Reference number: S48890; MUID:94088338; PMID:8264608
A:Accession: S48890
A:Molecule type: DNA
A:Residues: 1-257, 'L', 259-287, 'F', 289-707, 'L', 709-769 <PRO>
A:Cross-references: EMBL:U07187; NID:9460626; PIDD:AAA16835.1; PID:9460627
C:Genetics:
A:Gene: SGD:MLH1
A:Cross-references: SGD:S0004777; MIPS:YMR167W
A:Map position: 13R
C:Function:
A:Description: DNA repair
C:Superfamily: mismatch repair protein MLH1
C:Keywords: DNA repair

Query Match 29.6%; Score 1097; DB 2; Length 769;

Best Local Similarity 32.8%; Pred. No. 1.7e-59;
Matches 265; Conservative 156; Mismatches 244; Indels 144; Gaps 19;

```

OY 17 IRRLEESVYNNRIAGEVYQIORSASAVKELIENSIDAGASSVAVKDGKLIQVSDGCH 76
Db 4 RIKALDASVYNNRIAGEIITISPNALKEKEMENSIDANAMTIDLYVEGKIKVQIYTDNGS 63
OY 77 IRRFEDALICEHNTSKISAYEDLOTISKMGREGALASMTYVGHVYVTTTEGOLHGR 136
Db 64 GINKADLPILCEFTTSKIQKEDLSQIOTYGRGALASISVAVHTTTKADGCAVR 123
OY 137 VSYRGVNEKERPCAAVKGTOVMEVNFYNNVARKKTLONSNDYPRKYVDSISRAVHH 196
Db 124 RVSAYAGKMLSEPKPYAGKDGTTIYEDLFNIPSRALRSHNDYSKILDVYGRYAIH 183
OY 197 HINVTSCRHGANRADVSHASTSRDLAIRSYGASVVDLIEVYSYEDADSIFFKMG 255
Db 184 SKDIGFCKKFGDSNTSLVKSSTYQDRIYFNKSVASNLTFHISYEDL---NLESY 241
OY 256 DYTISNANYAKK-TMILFINDRLDCTALKRAIEFVYSATLPQASKPFYMSIHLPSE 314
Db 242 DGKVCNLTNISKSTISPIFFINRLVTCOLLRALMSYVSNITPKGNRPFTYIGIYIDRA 301
OY 315 HDVNIHPTKKEVSLNORIEETIRNAIEEKLMSNTTRIFOTQALNSGIAQANPOK 374
Db 302 ADVVNHPTKREVRFLSOEILIEKIANQHAELSAIDTSRTFASSISFTKKPSLIPND 361
OY 375 KV-----SEASMGSGTSQKIPVSOVMTDRPNPSGRILHT-----W 411
Db 362 TIESDRNRKSLROQVENSYTTANSOLRAKROENKLVARIDA--SQAKITFLSSSQOF 419
OY 412 HGOSSMLERKFPDVSRYNVRSSRRNOK-----DA-----GD- 442
Db 420 NREGSSTKQLSEPKTYNSHSEAKELTLNESEORPDANTINDNDLQPKKKQKGLGY 479
OY 443 -----SSRHELLVEIDSPHGLDIVKNCYV 471
Db 480 KYPSIADDEKNALPISKDGIYVPERVANNVNLISIKKLRKQVDSIHRRETDIFANLNY 539
OY 472 GLADE--AFALIOHNTRLYLIVNANNISKELMYQALCRFGNFNAIOLSEPAIOLLELM 529
Db 540 GVVDEERRLLAIOHDKLFLIDGVSICYELFYQIGLTPANFKIMQSTNVSDDIVLXN 599
OY 530 L-KODELSDKEDKLEIAEVNTEILKENAMINEYSIHIDOG-----KUTRL 579
Db 600 LLSPEDEL-NDQASKRTI-ISKI-----WDSMLNIEYSIELVNGDNDLDSVYLSL 652
OY 580 PVLIDQYTPMDRLPEFYALGNDVTWDEKECFRTVASAVNGFYALHPILPNSGNGI 639
Db 653 PLLKGIYIPSLVLPFIYIRKQEVWMEDEOCDIGILIREALTY---IP----- 699
OY 640 HLYKKNRDSMADEHAENDLSDENDVDOLLEAEAAQAQREWTI---QHVLFPMSRLPE 695

```

Db 700 -----DMPKVDTSASISEPKAOFINRKHISLSLEHVLPCIKRR 742

QY 696 LKPKSMATDGTFOVASLEKTKIFERC 724

Db 743 FLAPRHILKO--YVEIANLPDLTKVFERC 769

RESULT 5

probable DNA mismatch repair protein, MHL1 homolog [imported] - fission yeast (Schizosaccharomyces)

C:Species: Schizosaccharomyces pombe

C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 02-Sep-2000

C:Accession: T50317

R:McDougal, R.C.; Rajandream, M.A.; Barrell, B.G.; Cadieu, E.; Lelaue, V.; Galibert, F. submitted to the EMBL Data Library, January 2000

A:Reference number: 225061

A:Accession: T50317

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-684 <MCD>

A:Cross-references: EMBL:AL136536; PIDN:GAB6448.1; GSPDB:GN00067; SPDB:SPBC1703.04

A:Experimental source: strain 972h(-); cosmid cl703

C:Genetics:

A:Gene: SPDB:SPBC1703.04

A:Map position: 2

A:Insertions: 24/3; 70/3; 128/2

C:Superfamily: mismatch repair protein MHL1

Query Match 28.1%; Score 1040.5; DB 2; Length 684;

Best Local Similarity 32.7%; Pred. No. 4.3e-56;

Matches 241; Conservative 149; Mismatches 259; Indels 87; Gaps 13;

QY 17 RIRLESVYVRIAGEVIOEPSAVKELIENSIDAGASSVSAVAKDGLKIQVSDGCH 76

Db 8 KIRPDLQVINKIAAGEIIRPENAIKELIENSIDAGSTSIDVLLKDGKLLQITDMS 67

QY 77 GIRPDLALICERTTSKLSAYEDLOTIKSMGPRGALSMYGVHVTYITTEGOLHGY 136

Db 68 GIQYDDLPYLQORSTSKIDNFNDLQLOTFGRGALASISIVAKTYVTTLSSDIHM 127

QY 137 RVSIRGCV-----MENEPQCAAVKQVAVENTFYNNVARKTTLQNSNDYKTYDF 189

Db 128 KAFYVDALAPISPGMSPAPQPCGQGYITAEDEFYNNRSKSLKNGSEFRIMTL 187

QY 190 ISREAVHINVTSCRHGANRADYHASTSSRLDAIRSYGSAVYRDLEIVSYEDA 249

Db 128 KAFYVDALAPISPGMSPAPQPCGQGYITAEDEFYNNRSKSLKNGSEFRIMTL 187

QY 188 VOKAIHNDQVSEFKCKYGVDTVASLSLSSRLSKADKTRHITGPRVASHLDFSLGQSS 247

Db 250 DSIFKMOGYISNANVYAKKTMILFINDRLVDCALAKRAIEFYVSATLPQASKPIYMT 309

QY 248 IVGFANGFISNADQDKSMILFIFINRLVESVELHMLEYAKYLHKGASFYVLSL 307

Db 310 HLPSEHVNVNHTPKVEVSLNORLIETIRNAIEEKLANSNTIRIQOAL--NLGIA 367

QY 308 NNSPQDLVNVHPSKRIYHFLYDQELATISCDKLGELIERDTRSPLOMIPISINTK 367

Db 368 QANPOKDVSEKSGSKKIPVSOAVRTDPRNP--GLHTYMHGQSN-----LE 419

QY 368 NNE-----SSQAAVTYENTYLRTPRERSIKSLSONFLQSSNNNDNEIE 416

Db 420 KKFEDLVANVVRNRN-----QKADGISRHVELVIDSSFHGGLDIYKNCYV 471

QY 417 KYDSANSNKATNDIKDQTEIVEGNSIDLESIKSLQKQYINSMHVLAATINLEHKYV 476

Db 472 GLA--DEAFALQHNRLVAVNISKELVQALCFRGNFNAITOLSEPAPOELLYMA 529

QY 472 GLA--DEAFALQHNRLVAVNISKELVQALCFRGNFNAITOLSEPAPOELLYMA 529

Db 477 GIVCPTRIAAVOHNGLVVDYGLKSYHLFQICLTERGNGEVLEPLISIDLEIV 536

QY 530 LKDELSDEKDEKLEIAVNTIELKENAMINETYSHIDQDKLRLPVLDQVYD 589

Db 537 NGDE--DKSESEK-----TRILYSRDMKIDKFISIVSGGLTAVPMLSKYHP 586

QY 590 MDRLPEFVAL--GNDVTMDKEKCFRTVASAVGNFYALHPILPNSGNGIHLKKNRDS 648

Db 587 FEQPLLISSITKFPFMDKSCINGIMAKIF-----VPLP----- 626

QY 629 MADHAENDLISDENDVDOLAEANAKQREMTQIHLFSPMRFLKPKSMATDGT 708

Db 627 -----LSYSESDVKSIRSLES-----CLEDYLFEPFRNRVLCPKKVFBEKCI 668

QY 709 VQVASLEKTKIFERC 724

Db 669 YQITSLPRLVNVFERC 684

RESULT 6

hypothetical protein T28A8.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000

C:Accession: T25389

R:Lloyd, C. submitted to the EMBL Data Library, March 1997

A:Reference number: 220027

A:Accession: T25389

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-779 <MII>

A:Cross-references: EMBL:Z92813; PIDN:CA807283.1; GSPDB:GN00021; CESP:T28A8.7

A:Experimental source: clone T28A8

C:Genetics:

A:Gene: CESP:T28A8.7

A:Map position: 3

A:Insertions: 16/1; 77/3; 149/1; 205/2; 241/3; 404/1; 515/3; 696/3

C:Superfamily: mismatch repair protein MHL1

Query Match 22.8%; Score 845; DB 2; Length 779;

Best Local Similarity 29.7%; Pred. No. 5.3e-44;

Matches 240; Conservative 136; Mismatches 276; Indels 156; Gaps 22;

QY 18 RIRLESVYVRIAGEVIOEPSAVKELIENSIDAGASSVSAVAKDGLKIQVSDGCH 77

Db 25 IORLPQVNVNMAAGEVLNAPCAIKELIENSIDAGATELMVMQNGKLLQVSDNGK 84

QY 78 IRPFDLALICERTTSKLSAYEDLOTIKSMGPRGALSMYGVHVTYITTEGOLHGY 137

Db 85 IERDPLALYCFRSTSKLOKFEEDLMHMKTYGCFRGEALASISHAKVNIYSKRADAKAYQ 144

QY 138 VSYRDGVNENPKCAAVKQVAVENTFYNNVARKTTLQNSNDYKTYDFISREAVH 197

Db 145 ANFLDGKMTADTKPAAGKNGTCITATDFTNLPTRKMTTHGEBAKMNVDTLRFAIR 204

QY 198 INVFSCKRHGANRADYHASTSSRLDAIRSYGSAVYRDLEIVSYEDAADSIKMG 257

Db 205 PVSVALNQNCA--GDFRTKGDGNFRDVCNLLGRVADTILPLST--NSTRKFTFTG 259

QY 258 VTSN-----ANYAKKTMILFINDRLVDCALAKRAIEFYVSATLPQASKPIY 307

Db 260 HISKPIASATAIAONRTSRFSFVFTNGRSVRODILKHPIDVIGAR--QLHAQFCA 317

QY 308 SIHSEHVNVNHTPKVEVSLNORLIETIRNAIEEKLANSNTIRIQOALNLGIA 367

Db 318 HLOIDETRIDVNVHPSKRIYHFLYDQELATISCDKLGELIERDTRSPLOMIPISINTK 367

QY 318 HLOIDETRIDVNVHPSKRIYHFLYDQELATISCDKLGELIERDTRSPLOMIPISINTK 367

Db 368 QANP-----QKDVSEA-----SMGSGTSOKIPVSOAVRTDPR--- 401

QY 373 EQQPDIENTLWIMPQSLSKISIAIRKPDTPKPEKSPSAMSKDRDYMEVFRDAERK 432

Db 402 -----NPSGRL-----HTYMHGQSNLEKFPGLDVSVRNV-- 431

QY 433 IDEFVTRGCAVGTTSNDLIPGSSGLTKRARTEDSTGKEKEBEDINTDFDVSMSLST 492

Db 432 ---RSRRNQKACD-----LSRHVELVIDSSFHGGLDIYKNCYV 472

QY 493 ADGRILNESODLGEDDDVDVEYKTRHFRHPSIEVLRKELIANSQSILREKFKSTFVG 552

Db 410 IIKETSOEKEFYDITITTKAKKPELRY-----ICGFNNYTLIAESF 450
 Oy 485 TRLYVNVVNVISKELMYQOALCRFGNFAIQLSEPAFLDELLMALKDDELSDEKDK 544
 Db 451 EELVITIDQHAHEKLEFEKREDIKNG--VSQILITPSVELLPEDFIYDE-----502
 Oy 545 LEIAEVNTEILKENAEMENEYF--SIHIDODGKRLPVLLDQYPPMDRLPEFYLAIG 601
 Db 503 -----NNEVER-NAGEVIEFGDNTVAIKVPLFLGKPIYKDLFEIIDLKN-----MG 551
 Oy 602 NDVYDDEKECFRIVA--SAVGNFYAL 626
 Db 552 SGETSVKYRSIATAACSAVKAAYHEL 578

RESULT 9
 A95020
 DNA mismatch repair protein hexb [imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae
 C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
 C:Accession: A95020
 R:Jellel, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heide-
 on, J.D.; Umayam, L.A.; Holt, I.E.
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
 A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
 A:Reference number: A95000; MUID:21357209; PMID:11463916
 A:Accession: A95020
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-649 <KUR>
 A:Cross-references: GB:AE005672; PIDN:AAK74354.1; PID:g14971640; GSPDB:GN00164; TIGR:SP4
 A:Experimental source: strain TIGR4
 C:Genetics:
 A:Gene: SP0173
 C:Superfamily: mismatch repair protein hexb

Query Match 15.2%; Score 565; DB 2; Length 649;
 Best Local Similarity 27.7%; Pred. No. 6.2e-27;
 Matches 162; Conservative 115; Mismatches 219; Indels 88; Gaps 16;

Oy 18 IRLRESVNRVIAAGEVIOFPSAVKELIENSIDAGASSVAVKDGGLKIQVSDG 77
 Db 4 ITELPEMLANOIAGEVIERPAVKEVLENIDAGSSQIIEIEBAGIKKVOITDNG 63
 Oy 78 IREDELALICERTHTSKISAYEDLOTIKSMGFRGALASMTYVGHVYTTITTEBOLHGYR 137
 Db 64 IAHDEVELALRRHAKTSKIKNOADLFRITLGRGELALPSIASVSLITLTVADASHG 123
 Oy 138 VSYRDGVNENEPKCAVAGTQVAVENLFYNNVARKKTIONSDYPRKIVDFISRAVHH 197
 Db 124 LVARGGEVE-EVTPATSPGTRKVCEDLFENTPARLKYMSQOALSHIIDIVNRLGLAH 182
 Oy 198 INVTFSCKRKGANRADVHASSTSRDLAIRSVGASVVRDLIEIKVSTEDAADSTFKMDG 257
 Db 183 PELISFSLISDG--KEMTRTAGTGLRQIAIGIYGLVSAKKMIEIENS-----DIDFEISG 235
 Oy 258 YIS-----NANYAKKITMLFINDRLVDCALAKRAIEFVYATLPQASKPFIYMSIH 310
 Db 236 FVSLPELTPANRNYIS-----LFINGRYIKNFLNRNALIDGFGSKLMVGRPLAVYHH 289
 Oy 311 LPSEHDVNIHPTKEVSLINOERII-----ETIRNAIEE-----KLANSNTR 354
 Db 290 IDPYLADVNVHPTKQVRIKSKELMTLVSEALANSLEKQTLIPDALENLAKSTVNRREK 349
 Oy 355 IFQT-----QALNLSGIAOANPQKDYSEASMGSGIKSKIVISQVAVRDPNPSGRHL 408
 Db 350 VEOITLPLKENTLYEKEPESRPSQTEVADYVELTDEGDLTFAKETIDRLTKRAKHL 409
 Oy 409 TYWHGSSNLEK-----KFDLVSVRNVVRSRRNOKDAGDLSRHELLVEIDSSPHGLDI 464

Db 410 -FAERKPANYQJDHPDLIASI-----DKAYDKLEEREASPEL---449
 Oy 465 VKNCTYGLADEAFALIOHNTRLVYVNVVNVISKELMYQOALCRFGNFAIQLSEPAFLDE 524
 Db 450 -----EFFGQNHGTYLFAQGDGYIYIDQHAQBERVYEEYRESISNVDQSO-----QO 498
 Oy 525 LVNAL-----KDELSMDEK-----DDEKLEIAEV--NTEILKEN 558
 Db 499 LVPIYIEFPADALRLKENRPLLEEVGYLAIEYGNQFLEHH 542

RESULT 10
 H97891
 DNA mismatch repair protein [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
 C:Accession: H97891
 R:J. LeBlanc, D.J.; Lee, L.N.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.
 e, R.; Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.
 y, P.; Sun, P.M.; Winkler, M.E.
 J. Bacteriol. 183, 5709-5717, 2001
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
 A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
 A:Reference number: A97872; MUID:21429245; PMID:11544234
 A:Accession: H97891
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-649 <KUR>
 A:Cross-references: GB:AE007317; PIDN:AAK98964.1; PID:g15457702; GSPDB:GN00174
 C:Genetics:
 A:Gene: hexb
 C:Superfamily: mismatch repair protein hexb

Query Match 15.2%; Score 565; DB 2; Length 649;
 Best Local Similarity 27.7%; Pred. No. 6.2e-27;
 Matches 162; Conservative 115; Mismatches 219; Indels 88; Gaps 16;

Oy 18 IRLRESVNRVIAAGEVIOFPSAVKELIENSIDAGASSVAVKDGGLKIQVSDG 77
 Db 4 ITELPEMLANOIAGEVIERPAVKEVLENIDAGSSQIIEIEBAGIKKVOITDNG 63
 Oy 78 IREDELALICERTHTSKISAYEDLOTIKSMGFRGALASMTYVGHVYTTITTEBOLHGYR 137
 Db 124 LVARGGEVE-EVTPATSPGTRKVCEDLFENTPARLKYMSQOALSHIIDIVNRLGLAH 182
 Oy 198 INVTFSCKRKGANRADVHASSTSRDLAIRSVGASVVRDLIEIKVSTEDAADSTFKMDG 257
 Db 183 PELISFSLISDG--KEMTRTAGTGLRQIAIGIYGLVSAKKMIEIENS-----DIDFEISG 235
 Oy 258 YIS-----NANYAKKITMLFINDRLVDCALAKRAIEFVYATLPQASKPFIYMSIH 310
 Db 236 FVSLPELTPANRNYIS-----LFINGRYIKNFLNRNALIDGFGSKLMVGRPLAVYHH 289
 Oy 311 LPSEHDVNIHPTKEVSLINOERII-----ETIRNAIEE-----KLANSNTR 354
 Db 290 IDPYLADVNVHPTKQVRIKSKELMTLVSEALANSLEKQTLIPDALENLAKSTVNRREK 349
 Oy 355 IFQT-----QALNLSGIAOANPQKDYSEASMGSGIKSKIVISQVAVRDPNPSGRHL 408
 Db 410 -FAERKPANYQJDHPDLIASI-----DKAYDKLEEREASPEL---449
 Oy 465 VKNCTYGLADEAFALIOHNTRLVYVNVVNVISKELMYQOALCRFGNFAIQLSEPAFLDE 524
 Db 450 -----EFFGQNHGTYLFAQGDGYIYIDQHAQBERVYEEYRESISNVDQSO-----QO 498

QY 525 LVMAL-----KDELMASDEK-----DDEKLEIAEV--NTEILKEN 558
DB 499 LVPYIFERPADDALRLKERMPLEEVGFVLAEGENOFILREH 542
RESULT 11
Mismatch repair protein hexb [validated] - Streptococcus pneumoniae
C:Species: Streptococcus pneumoniae
C:Date: 27-Feb-1990 #sequence_revision 27-Feb-1990 #text_change 26-May-2000
C:Accession: A33589
R:Prudhomme, M.; Martin, B.; Mejean, V.; Claverys, J.P.
J. Bacteriol. 171, 5332-5338, 1969
A:Title: Nucleotide sequence of the Streptococcus pneumoniae hexb mismatch repair gene:
A:Reference number: A33589; MUID:90008767; PMID:2676973
A:Accession: A33589
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-649 <PRD>
A:Cross-References: GB:M29686; NID:9153656; PIDN:AAA88600.1; PID:9153657
C:Function:
A:Description: required for mismatch repair [validated, MUID:90008767]
C:Superfamily: mismatch repair protein hexb

Query Match 15.1%; Score 560; DB 2; Length 649;
Best Local Similarity 27.6%; Pred. No. 1.3e-26;
Matches 161; Conservative 115; Mismatches 220; Indels 88; Gaps 16;

QY 18 IRRLSESVNRIAGVIOEPSAVKELIENSIDAGASSVAVKGGKLIQVSDG 77
DB 4 IIEPLMALNOIAGVIERPASYCKELVENALDAGSSQIIEIEERGLKVOITDNG 63
QY 78 IRREDLAIICERHTTSKLSAYEDLOTIKSGFERGALASMTYVGHVYTTTTEGOLHGYR 137
DB 64 IAHDEVALLRRHATSKINKNOADLFRIITLGEALPSIASVYLTLTAVGASHGTR 123
QY 138 VSTRGCMENEPKCAAVGTOVMENLFYNNVARKKTLONSNDYPRKYDTSRPAVH 197
DB 124 LVARGGEVE-EVLPATSPVTKCVEDLFENFPAKLYKSOQELSHIIDYINRLGLAH 182
QY 198 INTFSCRHGANRADVHASTSSRLDAIRSYGASVNRLEIKVSYEDADSIIPKMD 257
DB 183 PEISFSLISDG--KEMTRAGTGOLROALAGILVSAKMIETENS-----DLDEISG 235
QY 258 YIS-----NANYAKITMILFINDRLVDTALRAIEFYSATLPQASKPFYMSIH 310
DB 236 FVSLPELFRANRYIS-----LFINGRYIKNFLNRAILIDGSGKLMGRPFLAYIH 289
QY 311 LPSEHVDVNIHPKKEVSLNDEKIT--ETRNAIEE-----KLMSNTTR 354
DB 290 IDPYLADVNVHPKQEVRSKELMTLVSEALANSIKEDTLIPDLLENLAKSTVRNREK 349
QY 355 IFOT-----QALNLSGIAOPKDKVSEASGSGTKSOKIPIVSOVWRTDPNPSGRH 408
DB 350 VEDTILPLAKENTLYEKTEPSRSQTEVADYOVLEDEGODTLFAKETIDRLTKRAK 409
QY 409 TYWHGSSNLEK-----KFLDVSVRNVRSRNOKDAGDLSRHLLVEIDSSPHGLDI 464
DB 410 -FAERKRPANVDQDHPDLASTI-----DKAVDKLEREASFPPEL--- 449
QY 465 VKNCITVGLADEAFALIOHNTRLIVNVNISKELMVOALCRGNFNAIOLSEPAIOL 524
DB 450 ---EFFGGMHGTLYLFAOGDGLYIIDHAAQERVYEEERESIGNVDSO-----OQ 498
QY 525 LVMAL-----KDELMASDEK-----DDEKLEIAEV--NTEILKEN 558
DB 499 LVPYIFERPADDALRLKERMPLEEVGFVLAEGENOFILREH 542

RESULT 12
H86900
DNA mismatch repair protein mutL [imported] - Lactococcus lactis subsp. lactis (strain
C:Species: Lactococcus lactis subsp. lactis

C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: H86900
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C:Accession: E64046
R:Flieschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kirlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodok, A.; Kelley, J.M.; Weidman
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Georgiagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Cross-References: GB:AE005176; PID:g12725273; PIDN:AAK06306.1; GSPDB:GN00146
A:Experimental source: strain IL1403
A:Genetics:
A:Gene: hexb
C:Superfamily: mismatch repair protein hexb

Query Match 15.0%; Score 555; DB 2; Length 695;
Best Local Similarity 30.7%; Pred. No. 2.8e-26;
Matches 178; Conservative 99; Mismatches 237; Indels 66; Gaps 17;

QY 17 RIRLESVNRRIAGVIOEPSAVKELIENSIDAGASSVAVKDGLKLIQVSDG 76
DB 42 KIIEALMALNOIAGVIERPASYCKELVENALDAGSSKIITINERAGRLLEVIDNGL 101
QY 77 GIRPEDLAIICERHTTSKLSAYEDLOTIKSGFERGALASMTYVGHVYTTTTEGOLHGYR 136
DB 102 GLEKEDEVALLRRHATSKINKNOADLFRIITLGEALPSIASVSOVWRTDPNPSGRH 161
QY 137 RVSTRGCMENEPKCAAVGTOVMENLFYNNVARKKTLONSNDYPRKYDTSRPAVH 196
DB 162 KLVAKGNET-LEPLAKRVGTISVANLEFNPAPKLYKSOQELSHIIDYINRLSLA 220
QY 197 HINVTSCRHGANRADVHASTSSRLDAIRSYGASVNDLIEIKVSYEDADSIIPKMD 256
DB 221 HPEISFLVNEG--KEMTRAGNGDLROVLAITIGSTAKMKQIK-----GSDLPELT 273
QY 257 GYIS-----NANYAKITMILFINDRLVDTALRAIEFYSATLPQASKPFYMSI 309
DB 274 GYSLPELFRANRYIT-----ILINGRPIKFLNRAILIDGSGKLMGRPFLAYIH 327
QY 310 HPESEHVDVNIHPKKEVSLNDEKITETRNAIEEKLMSNTTRIFOTOL- NLSGIAQ 368
DB 328 KIDPTLADVNVHPKQEVRSKELMTLVSEALANSIKEDTLIPDLLENLAKSTVRNREK 382
QY 369 ANPOKDKVSEASGSGTKSOKIPV-SOMVTRDPNPSGRHLYTYWHGSSN--LEKRP--- 422
DB 383 ---KEKES-LVQTEPLPIONNPLYDNRQDYVREAFKINKNOASDSEQTDFN 437
QY 423 -----DLVSVNRNVRSRNOKDAGDLSRHLL-VEID-----SSFHGLDIY 465
DB 438 TENQMTDAVSEKMTQRTVSTTEITDINSLENTVSNFEIDNEAKISQSTFEPOL--- 493
QY 466 KNTYVGLADEAFALIOHNTRLIVNVNISKELMVOALCRGNFNAIOLSEPAIOL 525
DB 494 ---EYLAQHLATYLLCCSGKGLYLVDOHAAQERKTEYWKDKIGEVSMEOIILAPY-L 548
QY 526 LVMALKDELMASDEKDEKLEIAEVNTEILKENAMIN 565
DB 549 FTLAKNDFYLAERKD--LLHAGVFLIEYGENOFILREH 566

RESULT 13
E64046
Mismatch repair protein mutL - Haemophilus influenzae (strain Rd KW20)

C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C:Accession: E64046
R:Flieschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kirlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodok, A.; Kelley, J.M.; Weidman
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Georgiagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A: Molecule type: DNA
A: Residue(s): 1-656 <PAR>
A: Cross-References: GB:AL162756; GB:AL157959; NID:97380091; PIDN:CAB84883.1; PID:9738
A: Experimental source: serogroup A, strain 22491
C: Geneticks:
A: Gene: hmuL; NMA1655
C: Superfamily: mismatch repair protein hexB
Query Match 14.5%; Score 538; DB 2; Length 658;
Best Local Similarity 25.7%; Pred. No. 2.9e-25; Indels 128; Gaps 20;
Matches 172; Conservative 122; Mismatches 246;

[illegible]

RESULT 15
 AH1612
 DNA mismatch repair protein [imported] - *Listeria innocua* (strain C1pil1262)
 C.Species: *Listeria innocua*
 C.Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
 C.Accession: AH1612
 R.Glaser, P.; Fraungul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
 R.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fshh
 D.; Jones, L.M.; Karst, U.
 Science 284, 849-855, 2001
 A.Mulhens, Krefl, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Maduno, E.; Maitouram, A
 Ok, G.; Schlutetel, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlh

A:Title: Comparative genomics of *Listeria* species.
 A:Reference number: AB1077; M01D:21537279; PMID:11679669
 A:Accession: AH1612
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-603 <GLA>
 A:Cross-references: GB:AL592022; PIDN:CAC96672.1; PID:g16413914; GSPDB:GND0178
 A:Experimental source: strain C11P11262
 C:Genetics:
 A:Gene: mult.
 C:Superfamily: mismatch repair protein hexB

Query Match 14.3%; Score 532; DB 2; Length 603;
 Best Local: Similarity 26.6%; Pred. No. 5.9e-25;
 Matches 173; Conservative 109; Mismatches 237; Indels 132; Gaps 19;

QY 18 IRLLESVVRIRIAGEVYQIPSSAVKELLINSIDAGASSVAVKDGKLIQVSDGHC 77
 DB 5 IVELTDLAKSKIAAGEVERPASYKELVEMALDAGSTVIDIYEAGLNKTIIDNGSG 64
 QY 78 IREEDLALICERHTTSKLSAYEDLQTKSMGFGEALASMTYVGHVVTITTEGOLHGYR 137
 DB 65 IEEEDVAIAFLRHATSKIKHEADLFRVHTLGFGEALPSIASVSHLLETST-GETKGT 123
 QY 138 VSTRDGMENEPKCAAVKCTQVAVENLFYNNARKKTIONSNDYPKIVDFISRAVHH 197
 DB 124 ILEGGKRI-TEOKSGHARKCTQIEVSOLFNTFARLRYLKLPLTELGNITDILNRLALAH 182
 QY 198 INTFSCRKHGANKRADVHASTSSRLDAIRSVGASVVRDLIEIKVSEDAADI-FKMD 256
 DB 183 PDISFRSHNG--KPLQITNGNDLROVIAITGVSTARKSIPK-----AESLDPKIS 234
 QY 257 GYI-----SNANYAKKITMILFINDRLDCTALRAIEFYVSATLPQASKPIYMSI 309
 DB 235 GYAVLPEVNSNRNYS-----TIINGREFIKPALVKAIOEGYHTLLPIGRPIIYLOI 288
 QY 310 HLPSEHVDVNIHPTKKEVSLINQERILETRNAIEEKLMSNTRIRIQTQALNLSGIAQA 369
 DB 289 EMPPIIVDVAVHPAKLEVRLSKEKEGLQISOMIKD-----AFHQQLIPDGEVSKK 340
 QY 370 NPQDKVSEASMGSGTFSOK-----IPVSQMV-----RT 398
 DB 341 QKEQOKSEQIQMSFEERKQPKPEPTLFSKNPIEYVSDEIVKREDDFILEMPTYNPET 400
 QY 399 DPRN-----PSGRILH-TYMHGQSSNLEKKEDLVSVRNVRSSRRNQKDGADLSS 445
 DB 401 OPEQVEOKERIPKMPYIGOMHATYIFAOENENGLYIIDQAAQERIKYEFYREKIGEVSR 460
 QY 446 R-HELLVEIDSSFRPGLDIVKNCITYGLDEAFALIQHNTRLYLNVNVISKELMTQQA 504
 DB 461 ELQELLVPIVLEPP-----SDEVYRLEROKAKLEEVGVF----- 494
 QY 505 LCRRGNFNAIQLESPAPLOELVWALKDDELMDEKDEKLEIAEVNTEILKENAEMI-- 562
 DB 495 LENFGONSFTIRAHPTWFK-----DQEEEMLERIIDEALSAISISIHKLREPTALMS 548
 QY 563 -----NEYFSIHIDQDGKLTRL-----PVVLDQYTPDMRL 593
 DB 549 CKKSIRKANHLYTMQ-DWEALDLTLREASDPFTCHGRPVIIQYSTVELEKK 598

Search completed: March 27, 2003, 15:45:36
 Job time: 29 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.D5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 15:42:15 ; Search time 42 Seconds

(Without alignments)
3551.860 Million cell updates/sec

Title: US-09-954-950-2

Perfect score: 3709

Sequence: 1 MDPSPRGGCGACGPEPRIR.....DGFVQVASLEKLYIFERC 724

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SP-archaea:*
2: SP-bacteria:*
3: SP-fungi:*
4: SP-human:*
5: SP-invertebrate:*
6: SP-mammal:*
7: SP-mhc:*
8: SP-organelle:*
9: SP-phage:*
10: SP-plant:*
11: SP-rodent:*
12: SP-virus:*
13: SP-vertebrate:*
14: SP-unclassified:*
15: SP-virus:*
16: SP-bacteriophage:*
17: SP-archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3645	98.3	715	10	O8S2D3	O8S2d3 oryza sativ
2	2505	67.5	727	10	O9C5F9	O9C5f9 arabidopsis
3	2505	67.5	737	10	O9ZRY4	O9Zry4 arabidopsis
4	1438	38.8	760	11	O8VDY4	O8vdy4 mus musculus
5	1261	34.0	664	5	O9V380	O9v380 drosophila
6	1255.5	33.9	663	5	O61917	O61917 drosophila
7	1040.5	28.1	684	3	O9P7W6	O9p7w6 schizosacch
8	911.5	24.6	887	5	O9B1X4	O9b1x4 trypanosoma
9	845	22.8	779	5	O9XU10	O9xu10 caenorhabdi
10	788	21.2	504	5	O8TON1	O8ton1 drosophila
11	684.5	18.5	484	11	O9CTA7	O9cta7 mus musculu
12	580.5	15.7	468	2	O9L4P6	O9l4p6 staphylococ
13	553	14.9	563	5	O8S500	O8s500 cephalalit
14	539.5	14.5	643	16	O8RG56	O8rg56 fusbacteri
15	530	14.3	590	16	O8RA70	O8ra70 thermomane
16	524.5	14.1	656	17	O8TBS5	O8tbs5 methanosarc

17	513.5	13.8	323	2	O93R03	O93R03 lactococcus
18	512.5	12.4	628	16	O9B0E0	O9B0E0 homo sapien
19	461	12.4	628	16	O8UH13	O8UH13 agrobacteri
20	429.5	11.6	923	10	O94116	O94116 arabidopsis
21	420	11.3	899	5	O9V7B6	O9V7B6 drosophila
22	414.5	11.2	908	3	O8RG50	O8RG50. saccharomyc
23	414.5	11.2	908	3	O8RG48	O8RG48. saccharomyc
24	414	11.2	895	5	O8T9C0	O8T9C0 drosophila
25	411	11.1	526	2	O9A155	O9A155 burkholderi
26	405	10.9	805	5	O9AVL8	O9AVL8 caenorhabdi
27	402	10.8	779	10	O8I287	O8I287 arabidopsis
28	401.5	10.8	908	3	O8RG57	O8RG57 saccharomyc
29	400.5	10.8	908	3	O8RG54	O8RG54 saccharomyc
30	388	10.5	893	5	O76417	O76417 drosophila
31	387.5	10.4	179	2	O93OH1	O93OH1 streptococ
32	354	9.5	380	3	P78862	P78862 schizosacch
33	325.5	8.8	788	5	O9BLV2	O9BLV2 trypanosoma
34	280.5	7.6	338	11	O8RA07	O8RA07 mus musculu
35	278	7.5	695	3	O07980	O07980 saccharomyc
36	277.5	7.5	1151	10	O8I785	O8I785 arabidopsis
37	276.5	7.5	389	4	O16530	O16530 homo sapien
38	266.5	7.2	252	4	O15157	O15157 homo sapien
39	266.5	7.2	715	3	O12083	O12083 saccharomyc
40	265	7.1	186	4	O16603	O16603 homo sapien
41	261	7.0	389	4	O13400	O13400 homo sapien
42	254	6.8	297	4	O95744	O95744 homo sapien
43	253	6.3	630	5	O8S500	O8S500 encephalito
44	232.5	6.3	166	4	O96H10	O96H10 homo sapien
45	231	6.2	119	3	O96V63	O96V63 magnaporthe

ALIGNMENTS

RESULT 1

ID	08S2D3	PRELIMINARY:	PRT:	715 AA.
AC	08S2D3:			
DT	01-JUN-2002 (TREMBLrel. 21, Created)			
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Putative MHL1.			
GN	P0401G10.8.			
OS	Oryza sativa (japonica cultivar-group).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;			
OC	Ehrhartoideae; Oryzeae; Oryza.			
OX	NCBI_TaxID=39947;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. NIPPONBARE;			
RA	Sasaki T., Matsumoto T., Yamamoto K.;			
RT	Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC			
RT	clone: P0401G10.8.			
RL	Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF003238; BAB89000.1;			
SQ	SEQUENCE 715 AA; 79943 MW; 75F99169C66F41AC CRC64;			

Query Match	98.3%	Score 3645;	DB 10;	Length 715;
Best Local Similarity	99.9%	Pred. No. 1.8e-233;		
Matches 712; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;
OY	1	MDPSPRGGCGACGPEPRIRLEESVVRNIAAGVIRPPSSAVKELIENSIDACASSVSA	60	
DB	1	MDPSPRGGCGACGPEPRIRLEESVVRNIAAGVIRPPSSAVKELIENSIDACASSVSA	60	
OY	61	VKGGGKLIQVSDGGRIFEDAIICERTTSKLSAYEDLQTKSMGFEGLASMTYV	120	
DB	61	VKGGGKLIQVSDGGRIFEDAIICERTTSKLSAYEDLQTKSMGFEGLASMTYV	120	
OY	121	GHTVTTTTEGOLAGRVSRDGVMEPEPCAAVVGCTQVWVNLVYMWARKKTQNSN	180	
DB	121	GHTVTTTTEGOLAGRVSRDGVMEPEPCAAVVGCTQVWVNLVYMWARKKTQNSN	180	
OY	121	GHTVTTTTEGOLAGRVSRDGVMEPEPCAAVVGCTQVWVNLVYMWARKKTQNSN	180	
DB	121	GHTVTTTTEGOLAGRVSRDGVMEPEPCAAVVGCTQVWVNLVYMWARKKTQNSN	180	

181 DDYKIVDFISRFVHHIINTFSCRHGANRADVHASTSSRLDARSVGYASVVDLIE 240
 181 DDIKIVDFISRFVHHIINTFSCRHGANRADVHASTSSRLDARSVGYASVVDLIE 240
 241 IKVYEDAADSTIFKMDGYSANVAKKIMILFINDRLVDTALAKRAIEFYASATLPOA 300
 241 IKVYEDAADSTIFKMDGYSANVAKKIMILFINDRLVDTALAKRAIEFYASATLPOA 300
 301 SKPFYMSIHLPSEHVDNIHPTKKEVSLNOERIEETINALEEKLMSNTTRIFQOA 360
 301 SKPFYMSIHLPSEHVDNIHPTKKEVSLNOERIEETINALEEKLMSNTTRIFQOA 360
 361 LNSGIAONAPKDKVSEASMGSGTKSOKIPVSOMVETDPNPSGRHLYTHWGOSSNLEK 420
 361 LNSGIAONAPKDKVSEASMGSGTKSOKIPVSOMVETDPNPSGRHLYTHWGOSSNLEK 420
 421 KFDLVSVRNVNRSRRNOAGLSSRHELVIEDSSPHGLDLYKCTYVGLADEAFAL 480
 421 KFDLVSVRNVNRSRRNOAGLSSRHELVIEDSSPHGLDLYKCTYVGLADEAFAL 480
 481 IOHNTRLYLVNVIKELMVOALCRFGNFNAIOLSEPAILOELLVMAKDELMDSDEK 540
 481 IOHNTRLYLVNVIKELMVOALCRFGNFNAIOLSEPAILOELLVMAKDELMDSDEK 540
 541 DDEKLEIAEVTLEIKENAEINIEFISIHDOGKILRLPVVLDQYTPMDRLPEFVAL 600
 541 DDEKLEIAEVTLEIKENAEINIEFISIHDOGKILRLPVVLDQYTPMDRLPEFVAL 600
 601 GNDVYMDDEKCEFRVAVASVGNFYALHPPLPNPSGNGIHLKKNRDSMADEHENDLIS 660
 601 GNDVYMDDEKCEFRVAVASVGNFYALHPPLPNPSGNGIHLKKNRDSMADEHENDLIS 660
 661 DENDVDOELAEAEAMAORENTIOHVLFPSSKRLFLPKPKSMATDGTGFVAVAS 713
 661 DENDVDOELAEAEAMAORENTIOHVLFPSSKRLFLPKPKSMATDGTGFVAVAS 713

RESULT 2
 09C5F9 PRELIMINARY; PRT: 727 AA.
 AC 09C5F9: PRELIMINARY; PRT: 727 AA.
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
 DE Putative MHL protein.
 GN AT4G09140.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_Taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
 Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 Kamada A., Karlin Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
 Seki M., Shinn P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R.,
 Tholomus A.
 RA "Full length cDNA of gene AT4G09140 (GI:7267557)." ;
 RT Submitted (MAR-2001) to the EMBL/Genbank/DBD databases.
 RL EMBL: AF360278; AAK25988.1;
 DR HSP: P23367; JAKN
 DR InterPro: IPR003594; Anpbind_Repair.
 DR Pfam: PF01119; DNA_mts_repair.1.
 DR Pfam: PF02518; HATPase_C.1.
 DR SMART: SM00387; HATPase_C.1.
 DR TIGRfam: TIGR00585; multi.1.
 DR PROSITE: PS00058; DNA_MISMATCH_REPAIR.1; UNKNOWN.1.
 DR PROSITE: PS00058; DNA_MISMATCH_REPAIR.1; UNKNOWN.1.
 SQ SEQUENCE 727 AA; 81301 MW; FDS532373AEL238C CRC64;

Query Match 67.5%; Score 2505; DB 10; Length 727;
 Best Local Similarity 66.4%; Pred. No. 8.5e-158;
 Matches 482; Conservative 111; Mismatches 129; Indels 4; Gaps 3;
 2 DEPSGGCCAGGPPRRRLREESVNRIRIAGEVIOIPSSAVKELIENSIDGASSVAV 61
 3 EESPPATTVIREPPKIORLEESVNRIRIAGEVIOIPSSAVKELIENSIDGASSVAV 62
 63 KGGGKLIQVSDGHIREDLILCRHTSKLSAYEDLQITKSGREGALASMTYV 121
 63 KGGGKLIQVSDGHIREDLILCRHTSKLSAYEDLQITKSGREGALASMTYV 122
 122 HVTVTITGQIHGRVSVRDGVNENEPKCAAVKGTQVAVENLFTNMAARKKTQNSND 181
 122 HVTVTITGQIHGRVSVRDGVNENEPKCAAVKGTQVAVENLFTNMAARKKTQNSND 182
 123 HVTVTITGQIHGRVSVRDGVNENEPKCAAVKGTQVAVENLFTNMAARKKTQNSND 182
 182 DYKIVDFISRFVHHIINTFSCRHGANRADVHASTSSRLDARSVGYASVVDLIE 241
 182 DYKIVDFISRFVHHIINTFSCRHGANRADVHASTSSRLDARSVGYASVVDLIE 242
 241 IKVYEDAADSTIFKMDGYSANVAKKIMILFINDRLVDTALAKRAIEFYASATLPOA 301
 241 IKVYEDAADSTIFKMDGYSANVAKKIMILFINDRLVDTALAKRAIEFYASATLPOA 302
 301 SKPFYMSIHLPSEHVDNIHPTKKEVSLNOERIEETINALEEKLMSNTTRIFQOA 361
 301 SKPFYMSIHLPSEHVDNIHPTKKEVSLNOERIEETINALEEKLMSNTTRIFQOA 362
 361 LNSGIAONAPKDKVSEASMGSGTKSOKIPVSOMVETDPNPSGRHLYTHWGOSSNLEK 421
 361 LNSGIAONAPKDKVSEASMGSGTKSOKIPVSOMVETDPNPSGRHLYTHWGOSSNLEK 422
 421 KFDLVSVRNVNRSRRNOAGLSSRHELVIEDSSPHGLDLYKCTYVGLADEAFAL 480
 421 KFDLVSVRNVNRSRRNOAGLSSRHELVIEDSSPHGLDLYKCTYVGLADEAFAL 481
 481 IOHNTRLYLVNVIKELMVOALCRFGNFNAIOLSEPAILOELLVMAKDELMDSDEK 541
 481 IOHNTRLYLVNVIKELMVOALCRFGNFNAIOLSEPAILOELLVMAKDELMDSDEK 542
 541 DDEKLEIAEVTLEIKENAEINIEFISIHDOGKILRLPVVLDQYTPMDRLPEFVAL 601
 541 DDEKLEIAEVTLEIKENAEINIEFISIHDOGKILRLPVVLDQYTPMDRLPEFVAL 602
 601 GNDVYMDDEKCEFRVAVASVGNFYALHPPLPNPSGNGIHLKKNRDSMADEHENDLIS 661
 601 GNDVYMDDEKCEFRVAVASVGNFYALHPPLPNPSGNGIHLKKNRDSMADEHENDLIS 662
 661 DENDVDOELAEAEAMAORENTIOHVLFPSSKRLFLPKPKSMATDGTGFVAVAS 713
 661 DENDVDOELAEAEAMAORENTIOHVLFPSSKRLFLPKPKSMATDGTGFVAVAS 714
 719 KIFERC 724
 722 KIFERC 727

RESULT 3
 09ZRV4 PRELIMINARY; PRT: 737 AA.
 AC 09ZRV4: PRELIMINARY; PRT: 737 AA.
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
 DE MHL protein (fragment).
 GN AT4G09140.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_Taxid=3702;
 RN [1]

102(b)

```

RP      SEQUENCE FROM N.A.
RA      Horwich M., Hilpert M., Kunze R.;
RT      "AtmH1, an Arabidopsis homologue of the E. coli mult. DNA mismatch
RL      repair gene."
RN      Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RA      Spiegel L.A., Huang E.N., Nascimento L.U., de la Bastide M., Val D.M.,
RA      Preston R.R., Matero A., Shah R., O'Shaughnessy A., Rodriguez M.,
RA      Shekhar M., Schutz K., See L.H., Swaby I., Habermann K., Dedhia N.N.,
RA      Mewes H.W., Lemcke K., Mayer K.F.X.;
RL      Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A.
RA      EU Arabidopsis sequencing project;
RL      Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AJ012747; CAI0163.1;
DR      EMBL; AL161514; CAB8038.1; -.
DR      HSSP; P23367; 1BKX.
DR      InterPro; IPR003594; AtPbind_ATPase.
DR      InterPro; IPR002099; DNA_mis_repair.
DR      Pfam; PF01119; DNA_mis_repair.1.
DR      Pfam; PF02518; HATPase_C.1.
DR      TIGRfams; TIGR00585; mult.1.
DR      PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; UNKNOWN_1.
SO      NON_TER
SO      SEQUENCE 737 AA; 82364 MW; 0497793964CA3700 CRC64;

Query Match          67.5%; Score 2505; DB 10; Length 737;
Best Local Similarity 66.4%; Pred. No.8.7e-158;
Matches 482; Conservative 111; Mismatches 129; Indels 4; Gaps 3

QY      2 DEPSRGGGCGAGEPRPRIRLRLEESVVRNRIAGEVIGIOPSSAVKELIENSIDAGASSVSV 61
DB      13 EEESPATITVPREPKRIQRLSESVNRRIAGEVIGIOPSAVKELIENSIDADSSISVV 72
QY      62 KDGGKILQVSDDGIGITBEDALICERTITSKLSAYEDLOITKSMGFGEALASITYG 121
DB      73 KDGGKILQVSDDGIGIRREDPLICERTHTSKLKFEDLSGSGFGEALASITYYA 132
QY      122 HVTVTTTIGGOLHGGRVSRVDCGMENEPKCAAVVGTGVWVNLFPVNVVARKKTIONSD 181
DB      133 HVTVTTTIGGOLHGGRVSRVDCGMENEPKCAAVVGTGVWVNLFPVNVVARKKTIONSD 192
QY      182 DYPKIVDIISRAVHHINWTFSCRRKHGKANRADVHASTSRIDATRSYGVASVVDLIBI 241
DB      193 DYKIVDLISRAHHINWTFSCRRKHGKANRADVHASTSRIDATRSYGVASVVDLIBI 252
QY      242 KVSITDADSTFKMDGYSNNANYVAKKTIIMLFIIDRLVDTALKRAITFVYSAILPOAS 301
DB      253 EVSSCDSSGCGCTFDMGFIISNSVYAKKTIYLFINDRLVECSALKRAITVYAALTPKAS 312
QY      302 KPEIYMSIHLPEHADVNVHPTKREVSLLNOERITETINALEEKLMSNTTRIPOTAL 361
DB      313 KPEIYMSIHLPEHADVNVHPTKREVSLLNOERITETINALEEKLMSNTTRIPOTAL 372
QY      362 NLGSIQAQNPQADKXSEASMSGSTKSQKIPVSQWVRTDPRNDSGRLHTYWHGOSSNLEKK 421
DB      373 EYIQ-STLTSQKSDSPVSOQPSGQTKQKVPVKNVRYTSSDPDAGRHLAIOKPSGLPK 431
QY      432 FDLYS-VNNVYSRRNQNDAGDLSRHELIVEIDSSFRHGLLDIYKNCITYGLADEAFAL 480
DB      432 VSSLSVSSVSSVQRNPNKPEFADLSVQELIADVDCSCCHGMEIYVNCITYGVADVFA 491
QY      481 IOHNRLVLYNVVNIISKELMYOALCRGFNFNAIOLSEPAIOLLYVMAIKODEL--MSD 538
DB      482 VOYNHTLIANYVNIISKELMYOALCRGFNFNAIOLSEPAIOLLYVMAIKODEL--MSD 551
QY      539 EKKDKLEIAVTEVTEILKENADINEYSIHAIIDODGKILRLRVVLDIQQYTPMDRIPEFY 598
DB      552 TKDOLKERIAEMNTTELKKAEMLEBYFSEVHIDSSANISRLPYIIDQYTPMDRIPEFY 611
QY      599 ALGNDVTDDECECRITYAASVGNFYALHPILPNDSGNGIHLKKNRDMSADEAENDL 658

```

Dd	612	CUGNDVEDEKSCQGSAAIGNTYAMHPLLPSPSGDGIPTFKRGGSSOEKSDLEGN	671
Oy	659	ISDENVDDELAEFAAQAOREMTTOHYLFPSPMRLLFLPKPSMATDGTFFVOVASLEKLY	718
Dd	672	VDMEDNLDDGLDSADENMAAGREMSIQHLFPMRLFLKPASMASNGFFVAVASLEKLY	731
Oy	719	KIFERC 724	
Dd	732	KIFERC 737	
 RESULT 4 O8VD14 PRELIMINARY; PRT; 760 AA.			
ID	O8VD14		
AC	O8VD14:		
DT	01-MAR-2002 (TREMBLrel. 20, Created)		
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Hypothetical 84.7 kDa protein.		
OC	Mus musculus (Mouse).		
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCHI_Taxid=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Straussberg R.;		
RL	Submitted (JAN-2002) to the EMBL/Genbank/DDBJ databases.		
DR	EMBL; BC021815; AAA21815.1;		
DR	InterPro; IPR003594; Atgbind_ATPase.		
DR	InterPro; IPR002099; DNA_mis_repair.		
DR	Pfam; PF01119; DNA_mis_repair.1.		
DR	Pfam; PF02518; NATPase_C.1.		
DR	TIGRFAMS; TIGR00585; mutl.1.		
DR	PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; UNKNOWN_1.		
SR	Hypothetical protein.		
SO	Sequence 760 AA; 84693 MW; AAA69BD99EF0F3B6 CRC64;		
 Query Match 38.8%; Score 1438; DB 11; Length 760; Best Local Similarity 38.5%; Pred. No. 5.9e-87; Matches 303; Conservative 153; Mismatches 216; Indels 112; Gaps 12			
Oy	18	IRLEEVSNNRIAGEVIQIPSSAYVELIENSIDAGASSVAAYKDGLKLIOVSDDGHG 77	
Dd	8	IRRLDETYYNRIAAGEVQIORPANAIKEMIENDCLDAKSTNIQVVAKRGKGLIQIDONGG 67	
Oy	78	IREFDLATICEHTSKLSAYEDLIOTIKSMGRGEALAMTYVGHTVTYTTEGOLAGR 137	
Dd	68	IKKEDIIDIVCEFTTSKLOTFEDLASISTYFGREGELASISHAHVTTITTKADGCARR 127	
Oy	138	VSYRDGVMEENERPCPAAVVGTQVMVENLFYNNAVARAKKTQNSNDVPKIVDFISPAVHH 197	
Dd	128	AAYSQGIQLQAPRKPCANGTTLITYEDLFYNNITRRKALKNSEEGKILLEVAGRSIH 187	
Oy	198	INVTYSCKRHGANRADVSHASTSSRLDAIRSYGASVVDLIEIKYSYDADSITFKMG 257	
Dd	188	SGSISISYKKQGETSVDRPLPNATYVDNTRISIFGNASRELLEVCGEDCTLA---FKMG 244	
Oy	258	YISNNNYVAKKLTMLTFLINDRLVDCAALKAILEPVYSATLPQASKFIWSTHLPSEHD 317	
Dd	245	YISNNAYSKKCIETLFEINHRLESAAALKALETYAAALPKTHPELTLSLEISPOND 304	
Oy	318	VNIHPFKKEVSLNOERILETTRNALIEEKLMSNTFRITOTQALNLSGIA----QANPQ 372	
Dd	305	VAVHPLKHVHFLHESILQRYQHIESKLILGSSNSRMFTYQL-LPGLAGPSGEARPT 363	
Oy	373	KOKVSEDASKSGTK-----SOKI-----PVSQMVRTDPNPSPGRLLHTYWG-- 413	
Dd	364	TGVASSSTGSGDKYAVQMYVRTRDSHQKIDLAFILOVSSVLSPQODPAPVRCARTGESBP 423	
Oy	414	-----OSSNEKEKFDEVLSVANVR-----SRNQDADQ 442	
Dd	424	ERATREDEMIALPAPAEEAAESENER-SLMETSDTQOKAPATISSPGSSRRKRHHEDD 482	

[illegible]

DB 653 NLEPTLYKFERC 664

RESULT 6

061917 PRELIMINARY; PRT; 663 AA.

AC 061917: 01-AUG-1998 (TREMBLREL. 07, Created)

DT 01-AUG-1998 (TREMBLREL. 07, Last sequence update)

DE 01-JUN-2002 (TREMBLREL. 21, Last annotation update)

DE MUTL homolog

GN MLH1 OR CG11482.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CANTON S;

RA McKee B.D., Borromeo D.;

RL Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.

DR EMBL: AF068257; AAC19117.1; -

DR HSP: P23367; 18KN.

DR Flybase: FBgn0011659; Mhl1.

DR InterPro: IPR003594; ATPbind_ATPase.

DR InterPro: IPR002099; DNA_mis_repair.

DR Pfam: PF01119; DNA_mis_repair; 1.

DR Pfam: PF02518; HATPase_C; 1.

DR TIGRfam: TIGR00585; multi; 1.

DR PROSITE: PS00058; DNA_MISMATCH_REPAIR_1; UNKNOWN_1.

SQ SEQUENCE 663 AA; 75848 MW; 590A666B2DB980B CRC64;

Query Match 33.9%; Score 1255.5; DB 5; Length 663;

Best Local Similarity 36.7%; Pred. No. 6,2e-75;

Matches 269; Conservative 145; Mismatches 223; Indels 95; Gaps 11;

DB 14 EPPRIRLESYVNRITAGEVIOEPSAVKELIENSLDAGASSVAVKDGKLTQVSD 73

DB 6 IPEGVIRKLDDEVYVNRITAGEIIRIPANALKELENSLDAGOSTHIOGVAVAGKLTQVSD 65

DB 74 DGHGIFEDLALICERHTTSKLSAYEDLOTISMGREGALASMTYGVHTVTTTTEGOL 133

DB 66 NGGIGIRREDLAIYCERFTSKLTFREDLSQIAIFGREGELASISVHLSIQTKAKEK 125

DB 134 HGTVSYRDGVNENKPCAAVKGTOVWENLFYNNVARKKTIONSDPKIVDSIRP 193

DB 126 CGKATYADGKLOGPCAGNOGTICIEDLFYNNPQRORALRSPAEEFORLSEVLAK 185

DB 194 AVHHINVTSCRRKHGADRADVHSASTSSRLDAIRSVYAGSVVDLIEIKVSYEDADSI 253

DB 186 AVHNPRVGTFLRKQDGAQALRTPVASSSENRIITGAISKELEF-----SHNDEV 240

DB 254 KMDG--YISNANVAKITMILFINDRLVDCIALKRAIEFVSATLPQASKPTIYMSIHL 311

DB 241 KEFAECILITOVNSAKCMLEFINDRLVDCIALKRAIEFVSATLPQASKPTIYMSIHL 300

DB 312 PSEHVDVNIHPTKEVSLNORIIETIRNAIEEKLMSNTTTRIFOTQALNSGIAQANP 371

DB 301 PPNLDVNIHPTKEVSLNORIIETIRNAIEEKLMSNTTTRIFOTQALNSGIAQANP 355

DB 372 QKDVSEASGSGTSKQIPVSQWRTDP-----RNPGRHLTYWHGSSNL 418

DB 356 DDEDTQAD-----KTQRIYPRKAVRTDSEQKDLFLAPLVKSDGVSSSSQEASRLP 410

DB 419 EKKFDLVSVANVRNRNODAGDLSRHELVLEIDSPHGLDITVKNCTYGLADAP 478

DB 411 EESFRTYAAK-----KSREVR-----LSSVLDMRKVRGQSVQRLTKNLVYGVCDERR 462

DB 479 ALQHTRLIYLVAVNISKEIMYQALCRGNNAIQLSPAPLOELVVALADDELMSD 538

DB 463 ALFQHTRLIYLVAVNISKEIMYQALCRGNNAIQLSPAPLOELVVALADDELMSD 522

DB 539 EKDKLEIAEVNTEILKENAEMINEYFSIHIDQDKLTRLPVYLDOYTPMDRLPEFV 598

DB 523 PEDEKAEILADGAADILKKAPIMREYFGLRISDEGDMLESPFLHQHRCVAAHLEPYLL 582

DB 599 ALGNDVTWDEKCEKRTVSAVGNFYALHPILPNFSGNGIHLKKNRDSMADEHENDL 658

DB 583 RLATEVDWDEKCEKRTVSAVGNFYALHPILPNFSGNGIHLKKNRDSMADEHENDL 608

DB 659 ISDENVDDELAEADANAO-----REWTIOHVLFPMSRLFKPKMSATDGTFOVA 712

DB 609 -----AOLDMREGATAVFSRMTMEHVLPAPKKTLLPFR---IKQDIYEL 651

DB 713 SLEKLYKIFERC 724

DB 652 NLEPTLYKFERC 663

RESULT 7

09P7W6 PRELIMINARY; PRT; 684 AA.

AC 09P7W6: 01-OCT-2000 (TREMBLREL. 15, Created)

DT 01-OCT-2000 (TREMBLREL. 15, Last sequence update)

DE 01-JUN-2002 (TREMBLREL. 21, Last annotation update)

DE Putative DNA mismatch repair protein, Mhl1 homolog.

GN SPBC1703.04.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI_TaxID=4896;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=972H-;

RA McDougall R.C., Rajandream M.A., Barrell B.G., Cadieu E., Lelaure V.,

RA Gallibert F.;

RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.

DR EMBL: AL136536; CAB66448.1; -

DR HSP: P23367; 18KN.

DR InterPro: IPR003594; ATPbind_ATPase.

DR InterPro: IPR002099; DNA_mis_repair.

DR Pfam: PF01119; DNA_mis_repair; 1.

DR Pfam: PF02518; HATPase_C; 1.

DR TIGRfam: TIGR00585; multi; 1.

DR PROSITE: PS00058; DNA_MISMATCH_REPAIR_1; UNKNOWN_1.

SQ SEQUENCE 684 AA; 77253 MW; 7A59633B36E2944B CRC64;

Query Match 28.1%; Score 1040.5; DB 3; Length 684;

Best Local Similarity 32.7%; Pred. No. 1,2e-60;

Matches 241; Conservative 149; Mismatches 259; Indels 87; Gaps 13;

DB 17 RIRRLLESYVNRITAGEVIOEPSAVKELIENSLDAGASSVAVKDGKLTQVSDGH 76

DB 8 KIRPLDQVLINKIAAGEIIRPENAIKELIENSLDAGASSVAVKDGKLTQVSDGH 67

DB 77 GIFEEDLALICERHTTSKLSAYEDLOTISMGREGALASMTYGVHTVTTTTEGOLHV 136

DB 68 GIDYDLPYLCOFSTSKINDNDLOHLOTTFGREGELASISVHAKYVTKLSSDIHAW 127

DB 137 RVSYRQGV-----MENEPKCAAVKGTOVWENLFYNNVARKKTIONSDPKIVDP 189

DB 128 KATYVGGALAPISPGSPAPQPCAGNOGTITAEDELFYNNRKRKSAKNSSEFRRLMIL 167

DB 190 ISRPAVHHINVTSCRRKHGADRADVHSASTSSRLDAIRSVYAGSVVDLIEIKVSYEDAA 249

DB 188 VQRYAIHQVQSPNCKKVGQTVASLSLSRLSKADKIRHRYGPRVASHLDFSLGEOSS 247

DB 250 DSIEFKMDGYSNANVAKITMILFINDRLVDCIALKRAIEFVSATLPQASKPTIYMSI 309

DB 248 IVFSGAFISNADPODKSNLILFINNRLVESVELHHALEETYAKYLHNGASYFYLSL 307

DB 310 HDPSEHVDVNIHPTKEVSLNORIIETIRNAIEEKLMSNTTTRIFOTQAL--NLGIA 367

Db 308 NMSPEOLDVNHSPKRIHLYDQELATISCKLIGLETERTERSTYPLQAMPSISINTK 367
 OY 368 QANQOKKVSFASMGSGTCKIPVSGWRTDPRNS--GRHTYWHQOOSN-----LE 419
 Db 368 NAE-----SSQKAVREYENTVLTDRERSITSMISDNFLQSNMNYDNEITTE 416
 OY 420 KREDVSVNRYVRSRN-----QKADGLSSRHELVLEIDSSHPGLDLYVNCITYV 471
 Db 417 KYDSANSMNKNATNDIKDQTEIEVEBGNISIDLESIKLOKQVNSMVLATNLEHKYV 476
 OY 472 GLA--DEAFALIOHTRLVYVNVNISKELMCOALCFGCFNAIOLSEPAPOLELLVA 529
 Db 477 GLVCFPRRLAAVOHNGILVYDYGKLSYHLFYQCLTEFGNGEYFLEPDISIDFEIV 536
 OY 530 LKDDLSMDEKDEKLEIAVNTLEIKENAMINEYSIHIDQCKITRLPVILADOTPD 589
 Db 537 NGDE---DSSEKFT-----TRLVSRHMDLKDYSISVTSGLTAVPMLSPKYHP 586
 OY 590 MDRLPEFVAL--GNDVTWDEKCECFRTVASAVGNFYALHPPLPNPNSGNGIHLKKNRDS 648
 Db 587 FEQPLILISSILPKFEDMDKESCLNGIMKALAKFY-----VPLP-----626
 OY 649 MADEHAENDLISDENDVQDELLAEAPAMAOBEMTIOHVLFPMSRLFKPKSMATDGT 708
 Db 627 -----LSESDVKSINSLES-----CLEDYLFPEFRFRVVCPEKVEEKC 668
 OY 709 VOVASLEKLYKIFERC 724
 Db 669 YOITSLPRLYVFERC 684
 RESULT 8
 O9BIX4 PRELIMINARY: PRT: 887 AA.
 ID O9BIX4
 AC 09BIX4
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
 DE Mult.
 OS Trypanosoma brucei.
 OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 NC NCBL_Taxid=5691;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN:ITATL.2;
 RA Bell J.S., Harvey T.I., Barry J.D., McCulloch R.,
 RT "Cloning and characterization of two Muts homologs from Trypanosoma brucei."
 RT Submitted (Feb-2001) to the EMBL/Genbank/DBJ databases.
 RL EMBL: AF346620; AAK29067.1;
 DR HSSP: P23367; IRLN.
 DR InterPro: IPR003594; ATPbind_ATPase.
 DR InterPro: IPR002099; DNA_mis_repair.
 DR Pfam: PF02518; DNA_mis_repair; 1.
 DR Pfam: PF02518; HATPase_C; 1.
 DR SMART: SM00387; HATPase_C; 1.
 DR TIGRFAIR: TIGR00585; mutL; 1.
 DR PROSITE: P500058; DNA_MISMATCH_REPAIR_1; UNKNOWN_1.
 DR PROSITE: P500058; DNA_MISMATCH_REPAIR_1; UNKNOWN_1.
 SQ SEQUENCE 887 AA; 95863 MW; 345531FE86BB97 CRC64;
 Query Match 24.6%; Score 911.5; DB 5; Length 887;
 Best Local Similarity 28.1%; Pred. No. 6; 6e-52;
 Matches 258; Conservative 134; Mismatches 281; Indels 245; Gaps 23;

OY 138 VSRDGVNENEPKCAVKGSTQVMEINLFYNNVARKKLONSNDYKIVDFISREAVHH 197
 Db 124 CHYVDGNGNEEPKPCAGNPGCTIRAKMFLYNAVRRAFSRSECSVEDVYSRALAP 183
 OY 198 INVTFSCKH-----GANRADVHASTSSRLDAIRSVYASVYRLIETIKSYSD-----247
 Db 184 PSVAFSCRRSDGSVAGYATKNCYCEPPDQSNLANIRHOMGEVASRLCEKRTGESPSDN 243
 OY 248 -----AADSIFKMDGYISNANYAKITMILFINRLVDCALRAIEFYSA 295
 Db 243 CIPESVILATSGSGEGRFLITGYSDITLASRKYLCVFNANNLVSTAIRALDAVSG 303
 OY 296 TEPQASKPTVSHLPSEHVDVNIHPTKKEVSLNDOELIETIRNAIEKINSWTERI 355
 Db 302 VLVGRNPPTVLPVYPPRVDVNIHPTKKEVSLNDOELIETIRNAIEKINSWTERI 363
 OY 356 FOTQALN-----LSCIA-----QANPK 373
 Db 364 MDIOLSKAVMLGDRESQSNOPMOPHSSTSPFNPPLPTGAGGVAAVAPCSIVREPOR 423
 OY 370 D-----FVSEASMGSTKSKIPVS---Q 394
 Db 426 GALDAFVRRKPTAEGNDAPLSESIIEERAGGVADQVRSAGTSSISKSDDTATGQNFQ 483
 OY 427 VARTADGKAPSGLTCAASSPANTVTYDAASSGSGARGWSENGSTGLSMTPVLLDT 543
 Db 408 -----HTYWHQOOSNLKREPLNS-----426
 OY 544 TDEDEGEVEYTHHEFKKRRKEVQ--DVVSVIDNAGVCGAGSDYIAEDNKAASGDA 602
 Db 427 -VANN--VSRNOKADG--LSSRHELVLEIDSSHPGLLDIVKCTYGLADEAFALI 481
 OY 603 AVREVGADSGASQBERAGFLITSVTSINIRAGTSGTQASLFPONLAYGVALKGLFPA 662
 Db 482 OHMTRLYVNVNISKELMCOALCF-----GNFNAIOLSEPAPOLELLVALKRD--EL 535
 OY 663 QSGTTLVYDLSLVHRYVYQRIELMATPQLSFPQLSFEPIHLSLDFALQVOL 722
 Db 536 MSDEK--DDEKLEIAVNTLEIKENAMINEYSIHIDQCKITRLPVILADOTPD 708
 OY 723 PPSOKRADGSGSLSLGRRLCNMRYMLODYFAEISADGHLIPLSMGTSMPPLRA 834
 Db 593 LPEFVALGNDVTMD--DEKCECFRTVASAVG--NEFA--LHPLLPNSGNGIHLKKNRDS 648
 OY 783 VPLFTWRLAAEVPYNAGEIECFALAHIAETLYGVQLHSSWLPVYIKDGLR-----834
 Db 649 MADEHAENDLISDENDVQDELLAEAPAMAOBEMTIOHVLFP--SMRLEFLPKSMATD 706
 OY 835 -----ODDVP--FCDA-----IRFGLLPCATNSTFTVPCCALVDG 869
 Db 707 TFOVASLEKLYKIFERC 724
 OY 870 TVQAVSVDELKVFERC 887
 RESULT 9
 O9XU10 PRELIMINARY: PRT: 779 AA.
 ID O9XU10
 AC 09XU10
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
 DE T2848.7 protein.
 GN T2848.7
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 NC NCBL_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lloyd C.R.

RL Submitted (MAR-1997) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode *C.elegans*: A platform for
 investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL; 292813; CAB07283.1; -
 DR HSSP; P23367; 1BKN.
 DR InterPro: IPR003594; Arpbind Arpase.
 DR InterPro: IPR002099; DNA_mis_repair.
 DR Pfam: PF02518; DNA_mis_repair; 1.
 DR TIGRfams; TIGR00585; multi; 1.
 DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; UNKNOWN_1.
 SQ SEQUENCE 779 AA; 88104 MW; 05038F525330C64 CRC64;

 Query Match 22.8%; Score 845; DB 5; Length 779;
 Best Local Similarity 29.7%; Pred. No. 1.4e-47;
 Matches 240; Conservative 136; Mismatches 276; Indels 156; Gaps 22;

 QY 18 IRRLESYVNRJAGEVIRPSSAKELIENSIDAGASSVAVKDGKLIQVSDGHC 77
 DB 25 IORLPDVNRMAAGEVIRPCKAIKELVENSIDAGATIMVMONGKLLQVSDNGK 84
 QY 78 IFFEDALICERHTTSKLSAYEDLOTIKSMGFEGEALASMTYGVYVTTTTEGQLHGYR 137
 DB 85 IREDEPALTCEPRTNSLOKPFEDLMHKTYGFGELALSLSHAKNIYISKRADACANQ 144
 QY 138 VSTRGVMEKNEPCCAVGVTOVVENLTYNNYARKKTLQNSNDYPKIVDFISRAVHH 197
 DB 145 ANFLDCKMADTKPAKGGKGTCTATDLEYNLEPTRRKMTTHGEAKMNDLIRPAIRH 204
 QY 198 INTFSCRGKGNARADVHSASTSSRLDAIRSVYGASVVDLEIKYSIEDAASIRKMG 257
 DB 205 PVSFELRQNA--GDFRKGDGNFQVYVONLGRVADTITLST--NSTRLKTFYTG 259
 QY 258 YTSN-----ANYVAKITMILFINDRLVDTALRALEFYVSAFLPQASKEFIYM 307
 DB 260 HISKPIASATAIAQNRKTSRSFVSFINGRSVRCOILKHIDEVLCAR--QLHAQFAL 317
 QY 308 SIHPSBEHDVNIHPTKREYSLNORRIETITNAIEKLMASNTTRITOTOLANLSCIA 367
 DB 318 HQIDETRIDVNVHPKNSVIFLEKEIIEIR-AYPEVYIG---EIRGFALDVEKE 372
 QY 368 QANP-----OKDVSEA-----SMGSGKSKOKIPVSOQ-VETDTR--- 401
 DB 373 EEOPIEENLWIMPSQSLSEIAIRKPTKPEKSSPSAMKSDKKRVYDVEFTDAKERK 432
 QY 402 -----NPSGRLL-----HTYMHGQSSNLKKEKFDLVSVRVV-- 431
 DB 433 IDEFVTRGAGVGTTSNDLIFGSSGILKRAKRTDSTGGEKEPEDLWTDPDVSMVLSVTS 492
 QY 432 ---RSRRNOKDAGD-----LSSRHLELVEIDSSFFHGLLDIYKNCYV 472
 DB 493 ADCRRILNESODLGEDDDVDPEYKGTREFFHESIEVLKKEILANSQSILREMKTSFPG 552
 QY 473 LADDEALILQHNRLYLAVNVNISKELMQALCRGFRNALDLSBP---LOELIYV 538
 DB 553 SINVKVLIQFGKSLYHDEFTYLRFFQIYSFSGANGSYRLDEPPAILIIELEL 610
 QY 529 ALKDELMSDEKDEKLEI-AEYNT-----EILKENAMINNEYFSIHIO--DGL--TR 578
 DB 611 ---GELSTREPNTAAFEYFANYENRFAAKELAEHADLHDFALKLDELGRHLHTE 666
 QY 579 LPVVLIDQYTPDMORLPEFVALAGNDYTWDEKECFRTVASAVGNFYALHPILPNPSGNG 638
 DB 667 IPELVAYFVQLEKLEPLLATVLANVDYDEQNTFTICAIIDLFTL----- 714
 QY 639 IHIYKKNRSMADENHENDLISDEN--DVDQELLAEAMAMQRENTIQHYVLEPSSRL 696
 DB 715 -----DTNFTITLDRKISAFSAIPW---KTLIKVLEMLPVKKRF 749

QY 697 KPPKSMATDGTFOVASLEKLYKIFERC 724
 DB 750 IPEHFKQAGVIRQLADSHDLYKVFERC 777

 RESULT 10
 ID Q8TON1 PRELIMINARY; PRT; 504 AA.
 AC Q8TON1;
 DT 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE GH181/P.
 GN MLH1.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA Stappleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuno J., Pacle J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celisner S.,
 RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AT069160; AAL39305.1; -
 SQ SEQUENCE 504 AA; 58261 MW; 9F6CF356F70066FC CRC64;

 Query Match 21.2%; Score 788; DB 5; Length 504;
 Best Local Similarity 31.4%; Pred. No. 4.4e-44;
 Matches 181; Conservative 111; Mismatches 191; Indels 94; Gaps 11;

 QY 169 MVARKKTQNSNDYKIVDFISRAVHHINTFSCRGKGNARADVHSASTSSRLDAIRS 228
 DB 1 MPOROALSPAEFEORLEVLARVAVHNPRVGFLLKQGDAPALTRPVASRSSENI 60
 QY 229 VVGASVVRDLEIKVGEDAADIIFKMDG--YISNNYVAKKITMILFINDRLVDCIALK 286
 DB 61 IYGAALSKELLEF-----SHRDEVYKFEACLLTYQVNSAKKCOMLFIQRLVSTALR 115
 QY 287 RATEFYVATLPOASKPEIYSHLPSBEHDVNIHPTKREYSLNORRIETITNAIEEK 346
 DB 116 TSVDSIYATILPGRHPFYMSLTPPQNDVNVHPKHEVHLYOBEIVDSIKQOYEAR 175
 QY 347 LKNSNTRTRIFOTQALNLSGIAQANPQKDVSEASMSGSTKSKPTVSOAVRTDP----- 400
 DB 176 LIGSNATRTFYKQ-LRLPGA-----PDLDEQLAD-----KTQILYKREWRTDSTQKLD 225
 QY 401 -----RNDGRLHYHMGQSSNLKKEKFDLVSVRVNRSRRNOKDAGLSSRHLEL 433
 DB 226 KFLAPLVKSDSGVSSSSQGRASRLPESFRTYAK--KSREYR-----LSSVLDMKRY 277
 QY 454 DSGFPHGLDIYVNCYVGLADAPALQHNNTLYLVNVNISKELMQALCRGFRNAL 513
 DB 278 EROCSVOLNSTLNLYVGVCDERRALFOHETRLMYCNTRSFSEELFYQMIYEFONCSE 337
 QY 514 IOLSEPAILOELLYMALKDDELMSDEKDEKLEIAEYNTIILKENAMINNEYFSIHIO 573
 DB 338 ITISPLPLKELILISLESAAQWTPBDGSKALADGADILKAPIMYEYGLRISD 397
 QY 574 GKILRLPVVLIDQYTPDMORLPEFVALAGNDYTWDEKECFRTVASAVGNFYALHPILPN 633
 DB 398 GMLSEPLSLHQHPCVANHLPVYLRLATEYVDEOETRCFETCRETARY----- 448
 QY 634 PSGNGIHLKKNRSMADENHENDLISDENVDQELLAEAMAMQRENTIQHYVLEPSSRL 697
 DB 449 -----AQLDMREGATGFSRWTHENY 469
 QY 688 LFPSSMLFLKPPKSMATDGTFOVASLEKLYKIFERC 724

DB 470 LPAKRYLLPERRIKDO--TYELTNLPITYKVERC 504

RESULT 11
ID 09CTA7 PRELIMINARY; PRT; 484 AA.

AC 09CTA7
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE 1110035C23R1K protein (fragment).
GN 1110035C23R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Eukaryota; Metazoa; Chordata; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
OX 11
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=EMBRIO;
RC MEDLINE=21085660; PubMed=11217851;
RX Kawai J., Shigaawa A., Shidara K., Konno H., Adachi J., Fukuda S.,
RA Aikawa K., Iwawa M., Nishi K., Kiyosawa H., Kasukawa T., Saito R.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Glasl C., King B., Kochava H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schmitt L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boileau D., Bojunga N., Carinici P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Holtman M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaurts P.,
RA Nordone P., Rong B., Schoenbach C., Rodriguez I., Sakamoto N.,
RA Sasakawa H., Sato K., Schoenbach C., Seiya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection."
RT Nature 409:685-690(2001).
DR EMBL: AK004105; BAB23172.1;
DR MGI: MGI:1915937; 1110035C23R1K.
DR InterPro: IPR002099; DNA_mis_repair.
DR Pfam: PF01119; DNA_mis_repair.1.
FT NON_TER 1
SQ SEQUENCE 484 AA; 54422 MW; 70DD1E60E2B1FAC5 CRC64;

Query Match 18.5%; Score 684.5; DB 11; Length 484;

Best local Similarity 30.8%; Pred. No. 3e-37; Mismatches 151; Indels 115; Gaps 11;
Matches 159; Conservative

DB 290 EHYATLPOAKRPTYSIHLPSEVDVNIHTPKKEVSLNORRIETIRNAIEKLM 349
DB 1 EHYATLPOAKRPTYSIHLPSEVDVNIHTPKKEVSLNORRIETIRNAIEKLM 349
DB 350 SNTRITCOALNLSGIA-----QANPKDKVSEASMSGSKTSOKIPVSOVFTD----- 399
DB 61 SNSSRMFTYOTL-LPGIAGSGEAPRTGVASSSTSSG--DKVAYQVWRDSEOK 116
DB 400 -----PNPNSGRILHTYWHG-----OSSNLEKK 421
DB 117 LDAFLQPVSLGPODPDPAVVGARTEGSPERATREDEEMALDAPAEAAESENLEKE 176
DB 422 PLYSVRNVVR-----SRNRKQAD-----LSSRH 447
DB 177 -SIMEISDAQAKAAPTSPGSKRRHRSDSVEMENASCKEMTACVRRRIINLTSL 235
DB 448 ELIVETDSFPGGLDIYKNCYVGLADEAFALIOHNTFLYLVNVNISKELMYOALOR 507
DB 236 SIOEELISERCHETREMLRNHSFVGVNOMALAOHOKTKLYLLNTTKLSLFTYQLIID 295
DB 508 FGNNFNAIOLSEPAIOLLYMALKDELMSDEKDEKLEIAEVNTEILKENAMINETS 567

DB 296 FANFVLRLESEPAIPEDRAMALDPSQWTEDDGPKGLAEVLYEELIKKAEMLADES 355
DB 568 ITHDQDKLTRELPAVLDQYTPDMRLPEFVLAIGDYTWDDKECFRTYASVANGFYALH 627
DB 356 VEIDEGMLGIPPLIDSYVPLEGPIPIFLRLATEVNNDEKECFEBSLSECMFYAIR 415
DB 628 PPLPNPSCNGIHLTKKNRDSMADEHAENDLISDENDVDOELLAEALAAQREKTOH 687
DB 416 KQYLE-----ESTLSGOQSDMP-----GSTSKFW--KWTYHI 447

DB 688 LPPSNMFLPKPKSMATDGTFOVASLEKLYKIFERC 724
DB 448 IYAFRSHLLPPEKFTEDGNVLOLANLPDLYKVERC 484

RESULT 12
ID 09LAP6 PRELIMINARY; PRT; 468 AA.

AC 09LAP6
DT 01-JUN-2000 (TREMblrel. 15, Created)
DT 01-JUN-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE DNA mismatch repair protein (fragment).
GN Staphylococcus aureus.
OS Staphylococcus aureus; Bacillus/Clostridium group; Bacillales;
OC Bacacteria; Firmicutes;
OC Staphylococcus.
OC NCBI_TaxID=1280;
OX 11
RN SEQUENCE FROM N.A.
RP STRAIN=COL;
RC MEDLINE=20031141; PubMed=10566865;
RX de Lencastre H., Wu S.W., Plinio M.G., Ludovice A.M., Filipe S.,
RA de Lencastre H., Sobral R., Gill S., Chung M., Tomasz A.;
RA Antibiotic resistance as a stress response: complete sequencing of a
RT large number of chromosomal loci in Staphylococcus aureus strain COL
RT that impact on the expression of resistance to methicillin."
RL MGI: MGI:1915937; 1110035C23R1K.
DR EMBL: AJ131755; CAB82461.1;
DR HSRP: P23367; 1IKN.
DR InterPro: IPR003594; ATPbind_Atpase.
DR InterPro: IPR004059; DNA_mis_repair.
DR InterPro: IPR004359; HIS_KIN_sig.
DR Pfam: PF01119; DNA_mis_repair.1.
DR Pfam: PF02518; HATase_C.1.
DR TIGRfam: TIGR00585; multi.1.
DR PROSITE: PS00058; DNA_MISMATCH_REPAIR_1; UNKNOWN.1.
FT NON_TER 468
SQ SEQUENCE 468 AA; 52987 MW; 87BADB2D50109FBE CRC64;

Query Match 15.7%; Score 580.5; DB 2; Length 468;

Best local Similarity 32.8%; Pred. No. 2.3e-30; Mismatches 145; Indels 69; Gaps 14;
Matches 149; Conservative

DB 17 RIRLESVYVNRRAAGEVORSSAVKELIENSUDAGASSVAVKDGCKLQVSDGH 76
DB 3 KIKELQTSIANKKTAAGEVERSSVYKLENAIDAGATEISIVESGVQSTRVNDNS 62
DB 77 GIREEDALICERTTSKLSAYEDLOTIKSMGFEALASMTYVGHVTTTTEGOLHG 136
DB 63 GIEAEDGLVYHRHATSKIDQDDEDLHRTIDGFRGALASISSVAKVLTCTD--NANG 121
DB 137 RVSITDQVMEVNEPKCAVAVGQVMEVFNKVAARKTLONSNDYKIVDFISRAVH 196
DB 122 ETVENGELIHL-KPAKAKGTDLIVSELYNPALKTKIKLYTELGLKIDYINRMA 180
DB 197 HINVTSCRHGANRADYHASTSSRDAIRSYGASVVDLIEIKVTEADAADSIKMD 256
DB 181 HPDIRIALISDCKTLSTNGSGRTN--EVAETIGKVAHDVHIS--GDTSD--YHIE 233
DB 257 GYV-----SNANVAKKITMLFINDRLVDCALRALEFYASATLPOAKRPTYSI 309


```

DB 234 GFVAKPEHSRKNHYS-----IFNGRYIKNEMLNKALLEGYHTLITIGREPICTINI 287
OY 310 HUPSEVVDVNIHPTKEVSLNORIEITIRNAIEEKLMSNTTRIFQOAL-----NLSG 365
DB 288 EMPDIIIVDVNHPTKEVSLNORIEITIRNAIEEKLMSNTTRIFQOAL-----NLSG 338
OY 366 IQAOPKDKVSEASMSGSKOKIPVSOQVTRDPNPSGRRLTYHGGSSNLEKFFDL 425
DB 339 V-----PKKKNVLSH-----PEOOKIEEORONTE-----NNOKTEFSSE 373
OY 426 SVANVYRRNOKDAGDLSRRHLEIVDISFPH 459
DB 374 ENSKRFEMENOND-----EIVAKED-SYNP 398

RESULT 13
OBS500 PRELIMINARY; PRT: 563 AA.
AC OBS500:
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE DNA mismatch repair protein (MUTL/HEXB family).
GN EC005_0300.
OS Encephalitozoon cuniculi.
OC Eukaryota; Microsporidia; Unikaryoniidae; Encephalitozoon.
OX NCBI_TaxId=6035;
RN RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RA Genoscope:
RL Submitted (Apr-2001) to the EMBL/Genbank/DBD databases.
RN RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RA MEDLINE=21576510; PubMed=11719806;
RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,
RA Prensier G., Barbe V., Peyretallade E., Brotlier P., Winkler P.,
RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
RA Weissenbach J., Vivares C.P.;
RT "Genome sequence and gene compaction of the eukaryote parasite
RT Encephalitozoon cuniculi."
RT Nature 414:450-453(2001).
DR EMBL: AL590445; CND25547.1;
SQ SEQUENCE 563 AA; 65296 MW; 266536004495B0F6 CRC64;

Query Match 14.9%; Score 553; DB 5; Length 563;
Best Local Similarity 24.2%; Pred. No. 2e-28;
Matches 174; Conservative 133; Mismatches 241; Indels 172; Gaps 22;

OY 18 IRRLEESVYVNRITAGVIOIPSSAVKELIENSLDAGASSVAVKDGKLIQVSDG 77
DB 3 IKRLPDAVIRISAGEVIRPYNLIKETIENSLDANSTHTTMEQDGLTL-IVEDDGS 61
OY 78 IREDALICERTITTSKLSAYEDLOTIKSMGFGEALASKTYGHVTV-TTTEGOLHG 136
DB 62 IHESDELLCKOYCTSKLKEELFSLSSYGEFGEALSSISRCARIKVSKRREGI-GY 120
OY 137 RVSIRGVVNEBPCAAVKGTOVMEVNFYNNVARKKTLQNSNDYPRKVIDISRPV 196
DB 121 EAVYRRTKEIT-ITKVGMDGIVETIKNTFYNNKVEKHSKRREIRREMLVGMISYF 179
OY 197 HINVTFSCKRHGANRADVHASTSSRLDAIRSYGASVVRDLIEIKVSYEDAASIFKMD 256
DB 180 NSRISFEL-FYGBKIQLELPKSRVCGEDGYSN-----EDRVKRVG--MLNEIYK 228
OY 257 G---YISNNYVA-----KTIIMLEFINDRLVDCALAKRAIEFYSATIPQASKPE 304
DB 229 GKILFVSDSEYLVITFSTOQFLKKGMLVLEVNERLAVSOEMKESLKYKDLIPKQPL 288
OY 305 IYMSIHLPSHVDVNIHPTKEVSLNORIEITIRNAIEEKLMSNTTRIFQOALNS 364
DB 289 IYELIYVSKMDVNVHPSKREVLFSNESMIOKLCIAERLSKID----- 335

```

```

OY 365 GIAOPKDKVSEASMSGSKOKIPVSOQVTRDPNPSGRRLTYHGGSSNLEKFFDL 424
DB 336 --YEOKPLK-PLKRSFQSPKIKVYSDPTSQSI-----AECLKE-- 371
OY 425 VSVRVYRRNOKDAGDLSRRHLEIVDISFPHGLDITVKNCTYGLADEAFALIQH 464
DB 372 -----TERRREFRLEFSKRLRLTYVDVDTF-----FRSLSYGVADRPTIIVOH 417
OY 485 TRLYVNVNISKELMYOALCRFGFNATQISEPAPLOELVWALKDDELMDEKDEK 544
DB 418 SLLNKCTVPLKEIKYLOSLINDGNF-----EKK 447
OY 545 LEIAEVNTEILKENAEMINEFYSIHIDQKTRLPVLDQYTPDMRLPEFYALGNDV 604
DB 448 RLVLRLCKMDKTRALINDFYSEIYEE-HYIGIPIL-----STICID-APF-LMSGFEI 500
OY 605 TWDEKECFRTVASAVNGFYALHPILPNPSGNGIHLYKKNRDSMADEHAEMLISD 664
DB 501 RRSSEYETLNIDVSTLY-----SGVEKSTKL-----FNI 532
OY 665 VDELLAEAAQAQREWTQHLFPMSRFLKPRKSMATDGFVQVYASLEKIKIFERC 724
DB 533 LKRRITGTARL-----ECFGLVTVTLKELYRNERC 563

RESULT 14
OBS56 PRELIMINARY; PRT: 643 AA.
AC OBS56:
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE DNA mismatch repair protein mult.
GN FN0462.
OS Fusbacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteriia; Fusobacterium.
OX NCBI_TaxId=76856;
RN RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RA MEDLINE=21866394; PubMed=11889109;
RA Bhattacharya V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goldsman E., Bernal A.,
RA Larsen N., D'Souza M., Malunas T., Pusch G., Haselkorn R.,
RA Fongstein M., Kyriades N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusbacterium
RT nucleatum strain ATCC 25586."
RT J. Bacteriol. 184:2005-2018(2002).
DR EMBL: AE010558; AAL94658.1;
SQ SEQUENCE 643 AA; 73742 MW; D7CE5D30E6D38770 CRC64;

Query Match 14.5%; Score 539.5; DB 16; Length 643;
Best Local Similarity 28.4%; Pred. No. 2e-27;
Matches 173; Conservative 109; Mismatches 211; Indels 117; Gaps 20;

OY 17 RIRLEESVYVNRITAGVIOIPSSAVKELIENSLDAGASSVAVKDGKLIQVSDG 76
DB 9 RIRLDESVNNAIAAGEVNEPPTSMIKELIENSLDAGSKIKLEVWNGG-RDISISDSGC 67
OY 77 GIFREDALICERTITTSKLSAYEDLOTIKSMGFGEALASMTYGVHVTVTTEGOLHG 136
DB 68 GMSKEDLLSIEBHATSKITRKEDLFNIRIYGRGEALSSIASAKMILSSRREDMONQ 127
OY 137 RVSIRGVVNEBPCAAVKGTOVMEVNFYNNVARKKTLQNSNDYPRKVIDISRPV 196
DB 128 QMNVVLGGKVTN-LKDIOKNVGTQIEIKDLFYNNPARKKFLKENTETYLNIKIDFLRAL 186
OY 197 HINVTFSCKRHGANRADVHASTSSR--LDAISYVGSVVRDLIEIKVSYEDAASIFK 254
DB 187 NPVAKFTLNIEGK-----ESIKTSGNGIENVALLEIFGKNYLNKFSKSL----- 230

```

Mon Apr 7 09:24:05 2003

us-09-954-950-2.rpt

25 MDGYSNAN-VAKKIMILFINDRLVDTALKAIEFVYSATLPQASKEPIYMSIHLP 313
OY 231 -GYLGNNLEKARNDISIPFINGRSYKSVKIEAVIAAHTKLMKGVYALFLEVEP 288
Db 231 -GYLGNNLEKARNDISIPFINGRSYKSVKIEAVIAAHTKLMKGVYALFLEVEP 288
OY 314 EHVAVNHPKREYSILNOELETETINALE-----EKINMSNTTRIFQ 347
Db 289 SEIDVAVHPSKRVKVEAQAQNAIFPLIGELNEFTDEDEFSIPYBAEVEENKKNF 347
OY 358 TQALNLSGIAQANPKQKVS-----ASWGSOTSKO-----IPVSQWVRDPNPSGRH 409
Db 348 -LDINDEFKDMODELQSLVYKGVYKSKYNNIKYKESFEDINK--KMT 395
OY 410 YMHGSS--NL-EKFEDLVSVNVRSSR-----NOKAGDLSRHLLAVE 452
Db 396 FGSAGTTESITLNEIKEDSKNIENFDSREINDKVKOTIFNOEDTG--RGLIFD 451
OY 453 IDSSFHGLDLYKNTY--VGLADEAFALIOHNTRLVYVNVNISKELMAYQALCHFGN 510
Db 453 IDSSFHGLDLYKNTY--VGLADEAFALIOHNTRLVYVNVNISKELMAYQALCHFGN 510
OY 452 FTS-----LKNIDFKVIGVDFTLIVERNGLEIYDQHIHERILYKLEKOEYN 502
Db 511 FNAIOLSPAPLOELV-----MALKDDELMS-----DEKDDKLEIAEYNT 552
OY 503 HSMK-----QNLVATREFELDPREKQALENIEFSSFGFDIDDEKKNILRSIPT 555
Db 553 ELIKENAVMI 562
OY 556 MNRDSYEMI 565

RESULT 15
Q8RA70 PRELIMINARY: PRT: 590 AA.

AC 08RA70; 2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE DNA mismatch repair enzyme (predicted ATPase).
GN MUTL OR TPE1358.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB47 / JCM11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao O., Tian Y., Li W., Xu Z., Huang L., Dong X., Ma Y., Ling L.,
RA Chen Y., Xue Y., Xu Y., Lai X., Yang H.;
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome."
RL Genome Res. 12:689-700(2002).
DR EMBL; AE013095; AAM24580.1;
KW Complete proteome.
SQ SEQUENCE 590 AA; 67196 MW; ASD01FB21A8A585E CRC64;

Query Match 14.3%; Score 530; DB 16; Length 590;
Best local similarity 25.7%; Pred. No. 7.4e-27;
Matches 149; Conservative 123; Mismatches 231; Indels 76; Gaps 14;

OY 17 RIRLEESVVRNAGEVIOBPSSAVKELIENSIDGASVSVAVDGGLKLIQVSDGH 76
Db 3 KIHLDKTVKNIAGEVVERPASIVKELVNSIDAGSKNIYVIELLEGIPYIKVDDGC 62
OY 77 GIRFEDALICERHTSKISAYEDLOTIKSMGFGEALASMTYGVHTVTTIEGOLHGY 136
Db 63 GMEIADAVLAFERHATSKIRSDDELFTITLIGREALASIAASKVYVLOTKEENETFG 122
OY 137 RVSRYDGMENEPKCAAAVGTQVWVENLFYVAVARKKTLQNSDDYPKIVDIPSPAVH 196
Db 123 RLVEGGLILEKTR-COCOGISVAVKDFPFTPARREKLRKPSTEAMYTEVETRLCLS 181
OY 197 HINTFGRKHGANRADVHSATSTSRDLAIRSYGASVVRDLIEIKVSEYEDNADSIKMD 256

Db 182 NPGISFYYVD--KVOFTISGNGSIEVILRLPKQEVHSAIIFSEFEADLKAKAFK 239
OY 257 GYISNANVAKKIMILFINDRLVDTALKAIEFVYSATLPQASKEPIYMSIHLPSEHV 316
Db 240 NPLVSN-----RNMKFYVNGRVKNTLSAAVDEAFKTYPSDRIGVLLYLEINPRT 295
OY 317 DVNIHPTKREYSILNOELETETINALE-----EKINMSNTTRIFQ 347
Db 296 DVNIHPTKREYSILNOELETETINALE-----EKINMSNTTRIFQ 347
OY 376 VGEASWGSOTSKOKIPVSQWVRDPNPSGRHITTYWHGSSNLEKFP-LVSVNVRSSR 378
Db 343 NEEGQIGEVK-LSLPFEV-----KEKTDGALFVREVEVT- 378
OY 435 RNOKAGDLSRHLLAVEIDSSFHGLDLYKNTYVGLADEAFALIOHNTRLVYVNVN 494
Db 379 -EKKIDRAPKHE-----SSDSERNVKKRLSDIRIYGLTFSTYVIEKGVFIIDQHA 430
OY 495 ISKELMAYQALCHFGNFAIOLSPAPLOELVYVNVNISKELMAYQALCHFGN 510
Db 430 AHERILYKLEKOEYN-----IVVELQPGD-----METVGGEREL 474
OY 555 LKENKEMINEFSIHIDODGKITRLPVLDTTPDMRL 593
Db 475 LYKLGYPEEFGNSV-----VIREVPVILGO--PEAKKL 507

Search completed: March 27, 2003, 15:45:06
Job time: 47 secs